

-2301-

5	Query: 539	FISHRFPGKANSQLNGNQTGSLFEPSVSKAPQSKGNMNNHPSNWGLTSDGGLKPDITAPG	598
	Sbjct: 494	FI RFGA++ + + F++ P+ + ++ PS+WGL++DG LKPD+ APG	549
10	Query: 599	GDIYSTYNDNHYGSGTSTMASPQAGASLLVKQYLEKTQPNLFKEKADIADIVKNLLMSNA	658
	Sbjct: 550	G IY+ NDN Y + GTSWASP +AGA++LVQYIL T P ++I +VK+LLMS A	609
15	Query: 659	QIRVNPETKTTTSPRQQAGLLMNDGAVTSGLYVTGKNYSGISLGNITDMTTFDVTVRN	718
	Sbjct: 610	+ HVN ET TSPRQQAG++ A+++GLY+TG-D YGSI+LGN+ DT +F VTN	669
20	Query: 719	LKNKDKTLRYDTLLTDHVDPQGRFTLTSHSLKTYQGGKVTVPANGKIVTVRVMVDSQF	778
	Sbjct: 670	++N+DKTL Y T+L ID + TS S +++ +VTV AN TV + +D S F	727
25	Query: 779	TKELTYKMPNGYLLSGFVFRFDSQDDQLNRVNI PFVGPKQFENLAVAEESIYRLKSGQK	838
	Sbjct: 728	+ELT M NGYLLSGFVEF D DD + V+IP+VGF+G+P+NLAV EE IY L +G K	786
30	Query: 839	TGFYFDE -SGPKDDIYVGKHTFLVTLGSETSNVSTKTSIDNGLHTLGTFFKNADGKFILE	897
	Sbjct: 787	GFYF+ + + + + H+TGLWT +E ST SD+ + TLGTFKN G P+L E	846
35	Query: 898	NAQNPVLAISPNGDNQDFAAFKGVFLRKYQGLKASVYHASNKHQPLAVS -PESFEG	956
	Sbjct: 847	+ G P LAISPNGD+NOD PKGVFLR Y L ASY A D E NPLM S +S G	904
40	Query: 957	DKN -FNSDIRFAKSTITLLGTAFSGKSLTAGELPDGHTHYVSSYPDVVGAKRQEMTFMI	1015
	Sbjct: 905	DKN ++ + + KS+ + T + + G G L DG Y YV++Y V GA Q M PD+I	964
45	Query: 1016	LDRQKPVLSQATFDPPEINRFKPELKDRLGAVKRDSSVFLERKDNKFPYTVTINDSYKV	1075
	Sbjct: 965	+DR+ PV++ AT+D F P P ++G +G+ ++ VFIL + T+ +	1024
50	Query: 1076	SVEENKTFVERQADGSFILPDKAKLGDFYMRVDRPAGNVAIAKL 1120	
	Sbjct: 1025	+V DNK PV + DGSF LVD A + FYY VED+AGH++ K+	1069

An alignment of the GAS and GBS proteins is shown below.

Identities = 543/1676 (32%), Positives = 821/1676 (48%), Gaps = 158/1676 (9%)

45	Query: 24	KQRFSLRKYKLGAVSVLLGTLFFLGGITNVAAD--SVINKSPDIAVQQVKDSTPSI---	78
	Sbjct: 4	KQRFSLRKYK G SVL+G++F + T VAAD S +++P+ QQ T+	62
50	Query: 79	ANETPTNNYSALATAQD-----NLVTKANNSPETQPVARESHQATETPSIVANQVPE	133
	Sbjct: 63	+ Be + +TS T ++ +LV++ + A + ++ A + P	122
55	Query: 134	STQEVSKPTLTKQ--NLAVESTPAISKETPNID-SNKIITVPKWNVTGYKGGTVPVAI-	189
	Sbjct: 123	+T +V TK + K + ID +++ + + V K + ++A	181
60	Query: 190	----IDSGLDIN-----HDALQINDSTKAK-----YQNEQNMNAKAKAGINTGKW	231
	Sbjct: 182	I+ G IN H+ ++ +D+ K ++N + A+ KA I K	240
65	Query: 232	YN-----NKVIFGHNYVDVNIELKEVKSTSHGMVTPSTATANPSKID-TNEL	277
	Sbjct: 241	Y + + G + +D + K SHGMVIT I N + T E	300
	Query: 278	TYGVAPEAQVMFMKVPFSDKRGKTPALYVKAIEDAVKLGADSNLSLGGANGSLVNADR	337
	Sbjct: 301	G+APEAQVMFMKVP+++ G+ +L+KATEDAV LGAD INLSL G ANG+ ++	360

-2302-

5	Query: 338	LKALEMARLAGVSVVIAAGNDGTQSGASKPSALYDPYGLVQSPSTAREAISVASYNBT	397
	Sbjct: 361	LMALEAKAKAGVSVVVAAGNRRVYSGDHDPLATNPDYGLVQSPSTGRTPTSVAALNSK	420
10	Query: 398	TLVNVFNIIIGLENNRNLNGLAAYA----DPKVSDDTKFVGGKQYDYVFGKGNDDYDKD	454
	Sbjct: 421	WVIGRLMTVKLELRADLNHGKALYSSGVDFDKDKSLGYDKSHQFAYKESDTAGYNAQ	480
15	Query: 455	TLANGKIALIERG-DITPTKKVVMNINHGAIVGAIIPNNKAGRANITMSLDPEASATPAIPT	513
	Sbjct: 481	DVIGKIALIERDPNKTICYEMIALAKHGALGVLPNNKPGQGNRSMRLTANGMGPASAFI	540
20	Query: 514	QKEPEDIAGKNYK----IVPNNIKNKQNPAGVLSDFSWSGLTDAGQIKPDLGAPGGS	569
	Sbjct: 541	SEFPGKMSQLNGKNTGSLFEDSVVSKAPSGKGNENHFSNGLTSDGYLKPDITAPGGD	600
25	Query: 570	IYAANDNEYDMMSTSMASPHVAGATALVKYLLKEHPALKKGDERTVKYLLMSTAKA	629
	Sbjct: 601	IYSTYNNHYQSQTSMASPGIAGASILLVKQYLEKTQPNLPEKEKADIKNLMSNAQI	660
30	Query: 630	HLAKDTGAYTSPRQGGAGIIDVAAAVQTGLYLITGGENYGSVTLGNIKDKISFDVTVHNI	689
	Sbjct: 661	HVNFEKTTTSPRQGGAGILLDGAIVTGLYVVG-KDNYGSLISLGNITDITMTFDVTVHNL	719
35	Query: 690	NKVKDLHYTYTNTDQV---KDGFTVLAPQGLFTPTGKTRIERPQSQQTITIDIVSKYH	747
	Sbjct: 720	SNKDKLRYDTELLDHDVDPQGRFTLTSLSKTYQGSEVTPVANGKTVRVIMVDSQPT	779
40	Query: 748	DMLAKVMFNGYLFYGRVFTDPVDSG-EVLSIPYVFGKGFENCLVLEKSIYKLVANKE	806
	Sbjct: 780	KLTKMFGNYLLEGFVRFRDSQDQLKNVNPFFVFGKGFENLVAEESLYRLKSGQKT	839
45	Query: 807	GFYFQPK-QTNEVPQSEDYITALMSTSEPIYSTDGTSPILKALGYSKSIDGKWLILQDQ	865
	Sbjct: 840	GFYF+++ +T L+T SE ST S L LG++K+ DSK+IL+ + GFYFDSGPKDDIYVGKHTGLVLTGSEINVTSTKISDNGLHTLTFKNADGKFLILEKA	899
50	Query: 866	KQPHLAISPNDDQNDQDAVAVKGVPLRPNFNRLAKVYRADDVNLKQPLMVSAQ-ACIDN	924
	Sbjct: 900	QCNFVLALSPGNDNNDFAAFKGVPLRKYQGLKASVYHSDKHNKPLNVSPSPKSDJN	959
55	Query: 925	YYSNTEMPKSTFLNDETEWKTITDGIPLDGKQKVLTYYSVDVPSKPKQMVDFLDR	984
	Sbjct: 960	FNS-DIRFAKSTTLGLTAFSGKSLTGAELPDGHTHYVVSVYTPDVVGAKQEMTFDMILDR	1018
60	Query: 985	QAPTLTATYDKDRIPKARPAVHEGSGSIPREBQVPLKDKDKDGHYSVLQGGEDGILV	1044
	Sbjct: 1019	QKFVLSQAFTDPETNRKPEPLIKORGLAGVKDSVFYLER-KONKQPTVTVINDSYKCVSV	1077
65	Query: 1045	EDNKVFIKQKDGSGFILPKEVNDPSHVYTVTRDAIGNLVSAKLELDINIGNKGLNVNKKV	1104
	Sbjct: 1078	EDNKTFVERQADSGFILPLDKAKLDGFYVHVEDFAGNVATKLDGHLPTQLKQTIPLKLK	1137
70	Query: 1105	FSPELMSNVDDIPSYSVKDDKGNILKKQ-----HSGKDLNLLKLPFGTYTFDLFLYDEE	1158
	Sbjct: 1138	TDGNYQTKETLKNLEMTQSDTGLVLINQQLAVHRRNQPSQLT---KMGQDFPISPN	1193
75	Query: 1159	RANLISPKSVTVTISEKDSKLDVLFKVNLLKKAALLVEPDKLLP-----KGATVQLVTKT	1213
	Sbjct: 1194	DGN----KDFVAFKLRNNVYNDL-TVNVYAKD---DHQKQTPINSSQAGASVATSEST	1244
80	Query: 1214	NTVVDLPKATYSPDYGKNI PUGDYRLNVTLPSCGYSTENLDDLNVSVKEDQVNLTL-KL	1271
	Sbjct: 1245	-----AWYGITARGSKVMPDQXQVTVYTRDEHG-ENQKQVTVSVNDKFMPTQGRFP	1295
85	Query: 1272	TLINK----APLINALABQTDITQOPVYRACHGLKONYLANLEAKQTLIKNRVSEQTSID	1327
	Sbjct: 1296	DTINGVDHFTPKTKALDSSGIVREKVFLA---KKNRKFQVTVSEKDKGI-----TVSD	1346

-2303-

Query: 1328 NATAALRESQALNGKSTTDTSLAKAILASTEIKGMYQFVNASPL----SQSTYIN---- 1379
 N + + + + D L+ + GN F L +N
 5 Sbjct: 1347 NKVYIPKNPDGSTTISRKDGVTLSDDYYLLVEDRAGHVSFATRLDKAVGDKKAVVNFGLD 1406

Query: 1380 -QVQLAKNLLQKPNVTQSVDKALEMLDIARNQLNGHETDYS--GLHHMIKANVLRQTS 1436
 V K ++ + + K +ENL+ N N Y + + N K S
 10 Sbjct: 1407 LPVPEDRQIVNFTYVREDAGKPIENLSEYNNNSGNSLILPYGKYTVELLTYDTNAAKLIS 1466

Query: 1437 SKYQNASQFAKENYNILIKKAEILLNSR-----QATQACQVRELLNLIKATSCQL- 1485
 K + + A N+ + K +L +++ + ++ ++ +Q+ BQ L
 15 Sbjct: 1467 DKIVSFLLSADNIFQQVTFKLTMLATSQITAHFDHLLPDSGRSLKTAQDQLTLEQSLY 1526

Query: 1486 -----DGRDKVSSARNYSQSLNNDNSLNTTFINPQNGQALIFKKMGTKES---- 1531
 +G V + + N +NT P N ++ + K G +S
 20 Sbjct: 1527 VPKAYGKTVQSGTYEVVVSLPKGYRIEGRYKVNLTLP-NEVHELSSLRLVKVGDAASDSTGH 1585

Query: 1532 -----EVAQKRVLGSVTSTNQKVKTNKLPKTBESTPKATITITILLFSLMLGLATI 1582
 +Q T LP TGE K+ + + L +LGL +
 25 Sbjct: 1586 KVMKNSQALATASATPTKSTTSATAKALPTGE---FMGLKRLVGLVLLGLTVC 1638

SEQ ID 8964 (GBS92) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 2; MW 48kDa).

GBS92-His was purified as shown in Figure 199, lane 9 .

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2038

A DNA sequence (GBSx2149) was identified in *S. agalactiae* <SEQ ID 6299> which encodes the amino acid sequence <SEQ ID 6300>. This protein is predicted to be AzlC family protein. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.80	Transmembrane	212 - 228 (196 - 230)
INTEGRAL	Likelihood = -7.27	Transmembrane	167 - 183 (159 - 185)
INTEGRAL	Likelihood = -5.68	Transmembrane	189 - 205 (188 - 210)
INTEGRAL	Likelihood = -2.28	Transmembrane	17 - 33 (13 - 34)
INTEGRAL	Likelihood = -1.06	Transmembrane	135 - 151 (135 - 151)
INTEGRAL	Likelihood = -1.01	Transmembrane	61 - 77 (60 - 77)

40 ----- Final Results -----

bacterial membrane	---	Certainty=0.4121 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

- 45 A related GBS nucleic acid sequence <SEQ ID 10235> which encodes amino acid sequence <SEQ ID 10236> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF01212 GB:AE001921 AzlC family protein [Deinococcus radiodurans]
 Identities = 72/224 (32%), Positives = 117/224 (52%), Gaps = 8/224 (3%)

50 Query: 6 FKEGVKDALPTALGYISIGLAPGIVASASDLSEVGLMSALVYGGSAQFAMCAILLAKA 65
 F +G + +P LG + LA+ + A A+ LS + LMS + G+QFA L A A
 Sbjct: 7 FWQGFRLVPLMLGTVPFALAYAVTARAGLSVGLTCLMSLTTFAGASQFAAGLPGHA 66

55 Query: 66 DLMTITNTVFLVNLRLNMLSLHATTIFKSAHLNMQLAGITLTDSEYGV-LLGALHKKV 124
 ++I +T FL+N R++L L + L ++ +TDE+YGV ++ A

-2304-

Sbjct: 67 GGLSIVITTFILNARHLLYGLSLARLRKIT-LPQRVVAQPIIDRAYGVAVVSGARLPGG 125

Query: 125 VSPSKWGRNNVMSYLTVVISTIIIGTLGSTIENPMPGLDFALVAMPFGLFVFLQFGLMS 184
 ++ +++ G + YL+W +ST++G L GS +D PE G+ F+GL V ++

5 Sbjct: 126 LTFAPLLGARELSLYLSMNWVSTLLGALAGSVLPPEQLGVGVVEPLAFPLGLIN----PLVV 181

Query: 185 DGRRLVVVYVLSVGLSYFLIATFLSGALSVLIATVVGCSVGTVVL 228
 D RL + V + GL + L+ L G L +LIA V G +G L

10 Sbjct: 182 D--RLSLVALAAGLGGWALSRLVPGELVILLAGVGGALLRAL 223

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2039

- 15 A DNA sequence (GBSx2150) was identified in *S.agalactiae* <SEQ ID 6301> which encodes the amino acid sequence <SEQ ID 6302>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

- 20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3794 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2040

- 30 A DNA sequence (GBSx2151) was identified in *S.agalactiae* <SEQ ID 6303> which encodes the amino acid sequence <SEQ ID 6304>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

- 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5087 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 40 A related GBS nucleic acid sequence <SEQ ID 10233> which encodes amino acid sequence <SEQ ID 10234> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:BA004157 GB:AP001508 homocysteine methyl transferase [Bacillus halodurans]
 Identities = 397/751 (52%), Positives = 519/751 (68%), Gaps = 14/751 (1%)
- 45 Query: 10 SNLGYPRIGEQREWKQATEAFWAGNLKQDKLEQLKQIRINHLKKQKEAGIDLIPVGDPS 69
 SNLGYPR+GE REWK+A+E+FWA + ++ L +K+LR+NLH+ Q+E +DLIPVGD+
 Sbjct: 4 SNLGYPRIGENREWKALLESFWANDTTEQLLATMKELRLNHLRVOQBSVDLIIPVGDFT 63
- 50 Query: 70 CYDHVLDLSPQFNVIPKRFDEY--ERNLDLYPALARGDKINVASSMKKWFNTNHYIYVPE 127
 YDHVLD++ F +IPKRF + L YFA+ARG K+ A M KW+STNYHYIYVPE

-2305-

Sbjct: 64 LYDHVLEMAVMGPIIKRFLQQGUTPTLSTYFAMARGSKNAQACBMKYNTNHYITVPE 123
 Query: 128 WEVETKPHLQNNYLLDLYLREAREVVGDKAKFVITGPITYVLSLSSGIVD--FEATVQRLLP 185
 + P L N L+ YLEA+ +G KPVI GP +V L+ G + +T+Q LLP
 Sbjct: 124 LH-DAAPRLTYNAFLSAYLREAKNELIGDKPVLIGPYSFKVLKAGYEEDKLQETIQSLLP 182
 Query: 186 LYKQVFDLIDAGATYIQIDEPIFVTDSEGLLVDAKSVYDFAREVPOAHFIPQTYFES 245
 LY QV Q+L+DAGA IQ+DEP VT + + +Y+ + A QTYF++
 Sbjct: 183 LYIQVIQELVDAGARSIQVDEPSLVTISAREMALVTRIYBQINEALADAPLFIQTYFDA 242
 Query: 246 AVCLDLKSLKLPVTGPGLODFIHKARNLAQVQ-GLFREKELFAGIVNGRNIWAVNLEETL 304
 +++ LDPV G GLOP+HG A+NL A++ G +K L AGI++GRNIW NL E
 Sbjct: 243 VTFEYEVSLPVGKIGLODFVHGGAKNLEALRTGFPEDKVLANGIDGNTIWSNLRERH 302
 Query: 305 ALLLEIGPFVK--RLTLQPSSESLHVPVTTKYETHLPDLVKNGLSPADEKLKLELLASA 362
 L+ ++ V EL LQPS SLHVPVTTK E LDP L L+PA+EKL EL L
 Sbjct: 303 ELVHQLSQAQVAKDRVLVQPSCSLHVPVTTKREKLOPTLLGVLPANEKLTSLHLEQL 362
 Query: 363 FQGNKTKGYHEALSR---PSALQAADPRHVALES-ABVKLERSPYKLRQALQAEKQL 417
 GN+ + EAL +AL+ +R A S E R + R+ LQ EK QL
 Sbjct: 363 AAGNEAB-VKRALEANDDAIAALEKSGWRSGAATSHNLENKKRQSPNERRPLQEEKVQL 421
 Query: 418 PILFTTTTISFPQSPSIRKKRLAWKGNLSDSDYKDFIKTEIRKRIAIQEDLOLOLVHVG 477
 P+LFTTTTISFPQ+ ++R R W+G LS +Y+ +K+ I +KI IQE+L LDVLVHVG
 Sbjct: 422 PLLFTTTTISFPQTYQVRRTRSLWRKGLSTVEYRTKMSYIEMKINIQEGLGLOLVHVG 481
 Query: 478 EFERVDMVEFPQGLKAGFTTTKLGWVQSYGSRVAKPFIYQDVKHQIPLSLESTVVAQSL 537
 EFER DMVEFPQ+KL GF T GWVQSYGER VKPFIYQ+V +P+++ STVVAQSL
 Sbjct: 482 EFERNDMVEFPQGLKGFAPTANGWVQSYGSRVAKPFIYQDVKHQIPLSLESTVVAQSL 541
 Query: 538 TKKPVKMLTGPITTIINWSEFDDISRSOLFNLALAIKORIQLLSGLIAIIVQDEAL 597
 T KPVKMLTGP+TI NWSF RDD+ + +QIA A+ E+ LE+GI +IQ+DE A+
 Sbjct: 542 TDRPVKMLTGPVTIINWSEFDDLFVIAHQAALTHEVLAEEGEMIGIQIDEAT 601
 Query: 598 REGLEPLRQQQAYLDVAARFKIATSSVKDETQINTHMCYSKFDEITDSIRALDADVIS 657
 REGLEPL+ QQ YLD AV+AF+ + + VK TQINTHMCYS+F E+I++I LDADVIS
 Sbjct: 602 REGLEPLRQEDQETLDMVSAFRAASHVAKATTQINTHMCYSFHEMIEALDLDADVIS 661
 Query: 658 IETSRSHGDIIESPETAVYPLGIGLVYDIHSPIPTKKEELIWNIGRSLSKCSKEQFVW 717
 IETSRSHG++I +FE Y IGLGVYDIHSPPR+P++E++ L+R+L L FVW
 Sbjct: 662 IETSRSHGEMISAFKTTYETKIGLVYDIHSFVPSSEEMLVIRRALTVLPASLFVW 721
 Query: 718 PDGSLKTRRAETIAALEVLVSATKEVRQQL 748
 PDGSLKTR E ET+AAL+ +VA+ + R++L
 Sbjct: 722 PDGSLKTRAKETVAALQWMAAARAAREEL 752

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2041

A DNA sequence (GBSx2152) was identified in *S.agalactiae* <SEQ ID 6305> which encodes the amino acid sequence <SEQ ID 6306>. This protein is predicted to be methH. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0753(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2306-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BA05340 GB:AP001512 unknown conserved protein [Bacillus halodurans]
  Identities = 301/610 (49%), Positives = 437/610 (71%), Gaps = 9/610 (1%)

5  Query: 1  MSKFLSKLKYDILLVADGAMETLLTYTGLTCHSSNVTHPEKVLAIHQAYIREAGADVIGT 60
    M+ +E LKT+ILV DGAMETILY G+D C E NVT PEK++A H AY+REAGADVIGT
    Sbjct: 1  NTNLV3ALKNTNLVGDGAMETILYEQGIDRCFSELAVTDPKTIVAHVAVYREAGADVIGT 60

10 Query: 61  NTYGAQRHLKLVGLEDQVVSINQAAVNIAHQATLKGHTFILGTGVSQFRSQQCDLTLN 120
    NTY A R +L Y L+DGV+ IN+AAV +A +A +ETF+LGT+GG RS + ++ +
    Sbjct: 61  NTYANRMKLAKYQLDDQVLEINRAAVLEARKAAK-QSTFVLGTIGIRSVQFEEVEIQE 119

15 Query: 121  IVESTLEQVEALLATGQLDGLLFETYDYIDIEETITVLKIVREMDLPITINISLHAGVTS 180
    + + LEQ++AI++ G +DGL ET+YD+EE + + R +TDLP+I ++S+ E GV
    Sbjct: 120  VQGVFLQKMLAVSEG-VDGLLETFYDLEERAKLAVSLARSLTDLPLVIAHLSIAEIGVLQ 178

20 Query: 181  NGKPTVEALSQVLM/GADVIGLNLCHLGPTHMIQSLKQVPLFAQSYLSVYNASQLIKGE 240
    GK + EA ++L LGAD++GANC +GPY H++SL+ V L ++Y S YHRS D
    Sbjct: 179  GKKLLEEFABLEGLGADLVGINCRMGPYQMLRSLETLVQLDRAYSGAYNASLP--DYR 236

25 Query: 241  NSQVQFSQNSSEYFKSABLLVAGVRLIGGCCGTTTPHIRAVKRSIRGLKPIERKVVTP 300
    + + + N EYF + + V +GVEL+GGCCGTTT+HARA + ++GLKP+ K V
    Sbjct: 237  DGRLYYHSNPEYFTFEMGRFVCGVRLIGGCCGTTTPEHVAFAKVVNKLKPVYSKPV-- 294

30 Query: 301  IPVKDFVRRIRRT---DILRVKVKETVILAELOPHHLDIVQFQKAIKIDQKGAIAIT 357
    ++K+ + + L +KVK+ ++II ELDPK+L I +F + A+ G+ A+T
    Sbjct: 295  LEIKETLSSTGQKAREFLAEKVKQPSIIVELDPKNIADRFVGAALKNAGVDAVT 354

35 Query: 358  LADNSLNTIRICNLISLASLKDSTPFLHITACRDHNLIGLQSLRLGELLGPNHILAI 417
    +ADNSL++ R+ NL++ ++++ ++ L+H+ CRD NLIGLQS L+G+ LQ +LAI
    Sbjct: 355  NADNSLASPRVDNALGAIITQQVGARPLVHVTCRDENLIGLQSHLMLHALGHTDLAI 414

40 Query: 418  TDGPTKLGDFPGATSVYDVISPKLLSLIKNLGLSVSGASLRPTDFTVAAPNPAVK 477
    TDGPTK+GDFPGATSVYDVIS+L+SLIKQLN+G+8+SG L + +F+V AANENV++
    Sbjct: 415  TDGPTKVGDFPGATSVYDVISPKLLSLIKNLGLSVSGASLRPTDFTVAAPNPAVK 474

45 Query: 478  LRTVKRLKLVKVASGADYFMTQPIPDHSLVKELADLTKTVEQPFPGIMPITSYNNAVFL 537
    L R V+ +EKK+ +GADYFMTQPI++ ++++ + TK +E+P +IGIMP+ + NA FL
    Sbjct: 475  LERAVQRMKKIKIENGADYFMTQPIYNEKQITEDIYATKHIEKPIYIGIMPLNORNAEFL 534

50 Query: 538  HNEVPGIKLSSESFLSALEKVKIDKKEACILTALNESKSLIDEALNYPNGIYLITPFLRVD 597
    HNEVPGIKL++ + + +D++ L +KSL+D A +YNGIYLITPFLRY +
    Sbjct: 535  HNEVPGIKLTDQIRERARAGEDRQKGEREGLAIAKSLLDVATHYPNGIYLITPFLRYGM 594

55 Query: 598  TLELDYIQK 607
    T+L Y+++
    Sbjct: 595  TVDLTHYVKE 604
```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2042

A DNA sequence (GBSx2153) was identified in *S.agalactiae* <SEQ ID 6307> which encodes the amino acid sequence <SEQ ID 6308>. Analysis of this protein sequence reveals the following:

```
55 Possible site: 53
>>> Seems to have no N-terminal signal sequence
  INTEGRAL Likelihood = -9.55 Transmembrane 127 - 143 ( 121 - 147)
  INTEGRAL Likelihood = -1.44 Transmembrane 157 - 173 ( 155 - 175)

60 ----- Final Results -----
      bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
```


-2308-

120 130 140 150 160 170 180

SEQ ID 8966 (GBS393) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 3; MW 30.8kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 177 (lane 4; MW 56kDa) and in Figure 83 (lane 6; MW 56kDa).

GBS393-GST was purified as shown in Figure 217, lane 5.

Example 2043

A DNA sequence (GBS2154) was identified in *S. agalactiae* <SEQ ID 6309> which encodes the amino acid sequence <SEQ ID 6310>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.29 Transmembrane 274 - 290 (271 - 291)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2317 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD35508 GB: A8001721 glycerol dehydrogenase [Thermotoga maritima]
 Identities = 94/307 (30%), Positives = 157/307 (50%), Gaps = 21/307 (6%)
 Query: 63 VYGDISTQSNIDKLKLVANPCVQAADAILGFGGKALDTAKMVAKEIGKNSPTIPTICSNCS 122
 ++G++ I++L ++ D++G GGGK LDTAK VA +L K +PTI S +
 Sbjct: 62 IFGEGCGDEEIERLGLVE-BETIVVVGIGGKTLDTAKAVAKLKKPVPVPTIASDTA 120
 Query: 123 AGTAAVVYNDHSLFLRYGY-PESPLHIFINTRIQAQPSKYPMAGIGDGSKAPVEERA 181
 +A+V+Y + F RY + P +P + +T I+A+AP++ AG+GD ++ E E
 Sbjct: 121 PCSALSVIYTFNGEFKRYLFLPRNPVVVLVDTEIVAKAPARFLVAGNGDALATWFEASC 180
 Query: 182 TLEAKINKLPH-TAVLQQAVALSSKEAFYQFGEGLKDVANLASRAVEEI--ALDILIS 238
 + N ++ A+A E ++G + VE + A+E+I A +L
 Sbjct: 181 KQKYAFNMTGRLGSMATAYALARLCYETLLEYGVLAKRSEKSVTPALEKIVEANTLISG 240
 Query: 239 TGVASNLVNPDPFYNSCHAHAFYGTITAIQRQGEFLHGVVAVFGLV-LHAYFNEISEL 297
 G+ S AHA + G T ++ ++LHG VA GVL L ++
 Sbjct: 241 LGFESG-----GLAAHAHINGLTVLENTHKYLHGKEVAIGVASLFLTDKPRNMI 291
 Query: 298 EKVARFNKSLGLFTTLADVSL---SEKDIPIKVEIAMTINE---YKNTPDFDMFQAAIL 351
 E+V F + +GLFTTLA++ L S++D+ K+ E A NE + P K A+
 Sbjct: 292 EKVYSFCEVGLFTTLARIGLDGVSDRLMKVAEKACDNETHNRFQFVTSKDVFFALK 351
 Query: 352 AADAFGQ 358
 AAD +G+
 Sbjct: 352 AADRYGR 358

There is also homology to SEQ ID 3078.

SEQ ID 6310 (GBS123) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 7; MW 43.3kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2309-

Example 2044

A DNA sequence (GBSx2155) was identified in *S. agalactiae* <SEQ ID 6311> which encodes the amino acid sequence <SEQ ID 6312>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0974 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6313> which encodes the amino acid sequence <SEQ ID 6314>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2368 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 92/167 (55%), Positives = 121/167 (72%)

Query: 1 MKIAIIGVSGSGKSTLARKLGNYNQVHLHDSIHFAFMWEERKYDDMIDVSNMLEKRT 60
      +KIAIIG+SGSGKSTLAR LG +Y+C V HLD +HF+ NW+ER DMI D+S L K+
Sbjct: 1 LKIAIIGHSGSGKSTLARFLGQHYHCEVFLDQLHFBSSWQERSDHDMLADLSTCLLQKD 60

Query: 61 WIIEGNYKKLLQGERLADAEIIFFDPNRFNCLWRAFKRYCKFRGKTRFDMA+GCPEKD 120
      IIEGNY LY+ER+++AD II+ +F+RF+C++RAFKRY +RGKTRFDMA+ C EK D
Sbjct: 61 LIIEGNYANCLYEERMSRADYIIIVNFRSFHCVYRAFKRYLVNFGKTRFDMA+DNCQEKFD 120

Query: 121 FEFISWILKDGSRNQLKKYQSVVQKYSKRTIVLTHQQLSHYMTI 167
      F+ WIL DGRS Q Y+ VV+ Y K +L +Q+ L Y+ +
Sbjct: 121 VAFVIMILLDGRSRNQLKKYQSVVQKYSKRTIVLTHQQLSHYMTI 167

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2045

A DNA sequence (GBSx2156) was identified in *S. agalactiae* <SEQ ID 6315> which encodes the amino acid sequence <SEQ ID 6316>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3874 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAAL1941 GB:X59250 initiation factor IF-1 [Lactococcus lactis]
Identities = 62/72 (86%), Positives = 70/72 (97%)

Query: 1 MAKEDVIEIGKVVTHPNAMPTVELENGHQIATVSGKIRKNYIRILGDRVTVENSPY 60
      MAK+DVIE++GKVV+THPNAMPTVELENGHQ+LAT+SGKIRKNYIRIL GD+V VE+SPY
Sbjct: 1 MAKDQVIEVDGKVVTHPNAMPTVELENGHQVIATISGKIRKNYIRILGDRVQVKSPEY 60

```

-2310-

Query: 61 DLTRGRITTYRKF 72
 DLTRGRITTYRKF
 Sbjct: 61 DLTRGRITTYRKF 72

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6317> which encodes the amino acid sequence <SEQ ID 6318>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.3253(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 67/67 (100%), Positives = 67/67 (100%)

Query: 6 VIEIEGKVVETMPNMFVLENGHQILATVSGKIRINRYIRILVGDRVTVMSPYDLTRG 65
 VIEIEGKVVETMPNMFVLENGHQILATVSGKIRINRYIRILVGDRVTVMSPYDLTRG
 Sbjct: 1 VIEIEGKVVETMPNMFVLENGHQILATVSGKIRINRYIRILVGDRVTVMSPYDLTRG 60

Query: 66 RITYRKF 72
 RITYRKF
 Sbjct: 61 RITYRKF 67

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2046

A DNA sequence (GBSx2157) was identified in *S.agalactiae* <SEQ ID 6319> which encodes the amino acid sequence <SEQ ID 6320>. This protein is predicted to be adenylate kinase (adk). Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAAM1940 GB:X59250 adenylate kinase [Lactococcus lactis]
 Identities = 146/214 (68%), Positives = 170/214 (79%), Gaps = 6/214 (2%)

Query: 1 MNLLIMGLPGAGKGTQAQAKVIEEFGVAHISTGDMFRAMNANQYTEMGRLAASYIDRGELVP 60
 MNLLIMGLPGAGKGTQA IV+ +GV HISTGDMFRAM N+TEMG+LAAS+IDRGELVP
 Sbjct: 1 MNLLIMGLPGAGKGTQAQEFIVKNGYVNHISTGDMFRAMNANQYTEMGRLAASFLDAGELVP 60

Query: 61 DEVYNGIVKERLAQDDIAEKGFLLDGYPTTIEQAHALDNTLEELGLRLDGVINIKVDPSG 120
 DEVYNGIVKERLA+DDI GFLLDGYPTT+QAHALD LEELG+LD V+NI V+P+
 Sbjct: 61 DEVYNGIVKERLAQDDIKASGFLLDGYPTTIDQAHALDTMLEELGLRLDGVNIVNPNII 120

Query: 121 LIERLGRINIRKRTGTFHKVFNPPV-----DYKSDYQQRQDKPTEVRRLDVNIAQ 174
 L++RLGR I R G T+HK+FNPD YQR DD PSTVK RLDVRI +
 Sbjct: 121 LVDRLSGRYICRNGATYHKIFNPPTKVBTCDVCGSHDLGYQRADVPETVKNRLDVNIKE 180

Query: 175 GEPILHYRKILGLVTDIEGNGEITVEFADVEKAL 208
 PI+EHY +IGLV +IEG QEI++V D++K L

-2311-

Sbjct: 181 SAPIIBHYTELGLVKMIBGBOQISQVTDIDKVL 214

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6321> which encodes the amino acid sequence <SEQ ID 6322>. Analysis of this protein sequence reveals the following:

5 Possible site: 17
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 208/212 (98%), Positives = 212/212 (99%)

Query: 1 MLLIMGLPGAGKSTQAKIVEEPGVAHISTGDMFRAMMANQTEMGRLAQSYIDKGELVP 60
 MLLIMGLPGAGKSTQAKIVEEPGVAHISTGDMFRAMMANQTEMGRLAQSYIDKGELVP
 Sbjct: 1 MLLIMGLPGAGKSTQAKIVEEPGVAHISTGDMFRAMMANQTEMGRLAQSYIDKGELVP 60

20 Query: 61 DEVINGIVKERLAEDDIAEKGFLLDGYPRITBQAHALDATALBELGLRLDGVNIKVDPS 120
 DEVINGIVKERLAEDDIAEKGFLLDGYPRITBQAHALDATALBELGLRLDGVNIKVDPS
 Sbjct: 61 DEVINGIVKERLAEDDIAEKGFLLDGYPRITBQAHALDATALBELGLRLDGVNIKVDPS 120

25 Query: 121 LIERLSGRIINRKIGETPHKVFNPVDYKEEDYQREDDKPKETVKRRLDVNAQGEPILE 180
 L+ERLSGRIINRKIGETPHKVFNPVDYKEEDYQREDDKPKETVKRRLDVNAQGEPILE
 Sbjct: 121 LVERLSGRIINRKIGETPHKVFNPVDYKEEDYQREDDKPKETVKRRLDVNAQGEPILE 180

30 Query: 181 HYRKLGLVTDIBGNQETVFPADVEKALLEK 212
 HYRKLGLVTDIBGNQETVFPADVEKALLEK
 Sbjct: 181 HYRKLGLVTDIBGNQETVFPADVEKALLEK 212

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 A related GBS gene <SEQ ID 8967> and protein <SEQ ID 8968> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 0
 MoG: Discrim Score: -1.04
 GVH: Signal Score (-7.5): -1.08
 Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 0 value: 6.79 threshold: 0.0
 PERIPHERAL Likelihood = 6.79 106
 modified ALOM score: -1.86

45 *** Reasoning Step: 3

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

over 213aa

55 Lactococcus lactis
 BGAD[8612] adenylate kinase Insert characterized
 SP[F27143] KAD_LACIA ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP, TRANSPHOSPHORYLASE). Edit
 characterized
 GP[44074] emb[CAM1940.1] XS9250 adenylate kinase Insert characterized
 PIR[S17987] S17987 adenylate kinase (EC 2.7.4.3) - subsp. lactis Insert characterized
 60 PIR[B44812] B44812 adenylate kinase (EC 2.7.4.3) - Insert characterized

[illegible]

SEQ ID 8968 (GBS114) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 9: MW 26.9kDa).

The GBS114-His fusion product was purified (Figure 108A; see also Figure 200, lane 8) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 108B), FACS (Figure 108C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Example 2047

A DNA sequence (GBSX2158) was identified in *S. agalactiae* <SEQ ID 6323> which encodes the amino acid sequence <SEQ ID 6324>. This protein is predicted to be preprotein translocase secY subunit (secY). Analysis of this protein sequence reveals the following:

Possible site: 35

```
>>> Seems to have an uncleavable N-term signal seq
```

50	INTEGRAL	Likelihood = -14.01	Transmembrane	217	-233	(209 - 240)
	INTEGRAL	Likelihood = -8.65	Transmembrane	314	-330	(307 - 334)
	INTEGRAL	Likelihood = -6.16	Transmembrane	369	-385	(363 - 392)
	INTEGRAL	Likelihood = -5.36	Transmembrane	19	-35	(17 - 40)
55	INTEGRAL	Likelihood = -3.93	Transmembrane	180	-196	(179 - 192)
	INTEGRAL	Likelihood = -3.93	Transmembrane	395	-411	(392 - 412)
	INTEGRAL	Likelihood = -2.56	Transmembrane	51	-67	(57 - 105)
	INTEGRAL	Likelihood = -2.02	Transmembrane	117	-133	(117 - 133)
	INTEGRAL	Likelihood = -0.64	Transmembrane	270	-286	(269 - 286)

----- Final Results -----

-2313-

bacterial membrane --- Certainty=0.6604 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9467> which encodes amino acid sequence <SEQ ID 9468> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GE:CAA41939 GB:X59250 SecY protein [Lactococcus lactis]
 Identities = 292/433 (67%), Positives = 361/433 (82%), Gaps = 2/433 (0%)

10 Query: 1 MFLKLRDALKVMVRNKILFTIFILLVFRIGTHITVPGINVKSLBQMGLPFLNMLNLV 60
 MF K L++A KVK VR +ILFTIFIL VFR+G HIT PG+NV++L+Q+ +LPFL+M+NLV
 Sbjct: 1 MFFKLRKEAFKVKQVRARILFTIFILVFRIGAHITAPGVNVNQLQOVADLPFLSMQLV 60

15 Query: 61 SGNAWRNPSVFSMGVSFYITASIVVQLQMDILPKFVBNQGEVGRKRLNQAIRYISLF 120
 SGNAWRN+S+F+MGVSFYITASI+VQLQMDILPKFVBN QGE+GRKRLNQAIRYI+L
 Sbjct: 61 SGNAWRNYSLFAMGVSPYITASIVVQLQMDILPKFVBNQGEVGRKRLNQAIRYITLV 120

20 Query: 121 LAFVCSIGITAGPNTLSSVALVKTPNVQTYLLIGAILITGSMVVTWLGQITDKGPGKGV 180
 LA QSIGITAGF +SS+ +V+ PN Q+YL+IG +LITGSMVVTW+GEOI +KGF+GV
 Sbjct: 121 LANAQSIGITAGFQAMSSLNIVQNPWQSYLMIGVLLITGSMVVTWNGEQINERKPGSGV 180

Query: 181 SMIIIPAGIISIPSAITTIYEDFFVNVRSIAITNSYIFVGILIVAVIAIVFFTFIQQAE 240
 S+IIPAGI+S IPSAI ++Y+ F+NV R S I S+IFV L++ + I++ TTF+QQA
 25 Sbjct: 181 SVIIPAGIVSGIPSAIKSVYDEKFLNVRESIPMSWIFVIGILISAIIVTYVTTFVQQA 240

Query: 241 YKIPQYTKLVQGAFTSSYLPLKVNPAVIVPFIASSITITPSTIIPFFQ--NGKRIM 298
 K+PIQYTKL QGAFTSSYLPL+VNPAGVIVPFI SITT P+TI+ F Q G + WL
 Sbjct: 241 RKVPIQYTKLVQGAFTSSYLPLRVNPAVIVPFIAGSITTAPATTIQLFLRQSGSIVGWL 300

30 Query: 299 TKLQELLNYQTPVGMIIYAILILFSFFYFVQVNPKEARNLQKNSYISIRPGRTE 358
 +LQ L+Y T GM+ YA+LI+LF+FFY+VQVNPKE ARNLQK SYIPS+RFG+ TE
 Sbjct: 301 STLQALSYTTWTGMLFTALLVLVLTFFYSFVQVNPKEARNLQKQGSYIPSVRPGKTE 360

35 Query: 359 EYMSLLKKLKATIGSVFLAFISLLPIAQAALHSSIALGTSLLILIAIGBGMKQL 418
 +Y+S LL +LAT+GS+FL IS++PI AQ L +ALGTSLLILI I++KQL
 Sbjct: 361 KYVSRLLKRLATVGSFLGLISIIPIACNVWGLPKTVALGTSLLILIQVAIQVQKQL 420

40 Query: 419 GYLLKRYVGFPM 431
 GYLLKRY GFPM
 Sbjct: 421 GYLLKRYAGFMD 433

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3987> which encodes the amino acid sequence <SEQ ID 3988>. Analysis of this protein sequence reveals the following:

45 Possible site: 55
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -14.70	Transmembrane	233 - 249 (226 - 255)
INTEGRAL	Likelihood = -8.12	Transmembrane	330 - 346 (323 - 350)
INTEGRAL	Likelihood = -6.10	Transmembrane	384 - 400 (378 - 403)
50 INTEGRAL	Likelihood = -5.20	Transmembrane	35 - 51 (33 - 56)
INTEGRAL	Likelihood = -4.09	Transmembrane	199 - 215 (195 - 215)
INTEGRAL	Likelihood = -3.56	Transmembrane	167 - 183 (165 - 184)
INTEGRAL	Likelihood = -1.65	Transmembrane	411 - 427 (411 - 428)
INTEGRAL	Likelihood = -1.49	Transmembrane	133 - 149 (133 - 149)
55 INTEGRAL	Likelihood = -0.64	Transmembrane	286 - 302 (285 - 302)

----- Final Results -----

bacterial membrane --- Certainty=0.6880 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 60 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2314-

Identities = 377/434 (86%), Positives = 417/434 (95%)

Query: 1 MFLKILRDALKVEMVRNKILFTIPILAVFRIGTHITVPGINVKSLBQMGLPFLMANLV 60
 Sbjct: 17 MFLKIL+DALK+K VRNKI FTIPIL+LVFRIGTHITVEG+N KSLAQ+ ELPLFLMANLV
 MFLKILDKALKITVRNKIFFTIPILVFRIGTHITVPGVNAKSLBQLSELPLFLMANLV 76

Query: 61 SGNAWRNFSVFSMGVSPYITASIVQLLQMDILPKFVEMGKQEVGRKRLNQATRYISLF 120
 Sbjct: 77 SGNAWRNFSVFSMGVSPYITASIVQLLQMDILPKFVEMGKQEVGRKRLNQATRYISLV 136

Query: 121 LAFVQSIGITAGPNTLSSVALVKTPNVQVITYLLIGAILTIGSMVVIVLGEQITDKGPGNGV 180
 LAF QSIGITAGPNTLS+VALVKTP++TYLLIGA+LITGS++VIVLGEQITDKGPGNGV
 Sbjct: 137 LAFQSIGITAGPNTLSNVALVKTPDKITKYLLIGAILTIGSVIVLGEQITDKGPGNGV 196

Query: 181 SMIIIPAGIISSIPSAITTIYEDFFVNRSSAITNSYIFVGILIVAVLAVFPFTTQQAPE 240
 SMIIIPAGIISSIPSAI TI ED+PVNV++S+ +SY+ VGILI+AVLAVFPFTT+QQAPE
 Sbjct: 197 SMIIIPAGIISSIPSAITIREDFVNRKASDLSSYILVIGILIAVLAVFPFTTQQAPE 256

Query: 241 YKIPIQYTKLVQAGTSSSYLPLKVNAGVIVPFIASSITTIPSTIIPFFQNGKEIPWLTK 300
 YKIPIQYTKL+QAGTSSSYLPLKVNAGVIVPFIASSITTIPSTIIPF QNG++PWL+
 Sbjct: 257 YKIPIQYTKLVQAGTSSSYLPLKVNAGVIVPFIASSITTIPSTIIPFQNGRDLFWLNK 316

Query: 301 LQELNLYQTPVGMIIYAILIILFSFFYTFVQVNPKEKTAENLQNSSYIPSIRPGRETEY 360
 LQ+ + NYQTPVGMILYA+LIIILFSFFYTFVQVNPKEKTAENLQNSSYIPS+KPGRETE++
 Sbjct: 317 LQEIFNYQTPVGMIVYAILIILFSFFYTFVQVNPKEKTAENLQNSSYIPSVRPGRETEQ 376

Query: 361 MSSLLKKLATIGSVFLAFISLLEPIAQAALHSSSIALGGTSLILIIATIGDKMKOLEGY 420
 MS+LLKKLAT+G++FLAFISL PI AQAAL+LSSSIALGGTSLILI+TGIEGKMKOLEGY
 Sbjct: 377 MSALLKKLATVGAIFLAFISLLEPIAQAALHSSSIALGGTSLILIIISTGIEGKMKOLEGY 436

Query: 421 LLKRRYVGFMMNTE 434
 LLKRR+YVGFMMNT E
 Sbjct: 437 LLKRRYVGFMMNTE 450

A related GBS gene <SEQ ID 8969> and protein <SEQ ID 8970> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 MoG: Discrim Score: 6.16
 GVH: Signal Score (-7.5): -4.32

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 9 value: -14.01 threshold: 0.0

Integral	Likelihood	Transmembrane	217 - 233 (209 - 240)
Integral	Likelihood = -9.98	Transmembrane	311 - 327 (307 - 334)
Integral	Likelihood = -6.16	Transmembrane	369 - 385 (363 - 392)
Integral	Likelihood = -5.36	Transmembrane	19 - 35 (17 - 40)
Integral	Likelihood = -3.93	Transmembrane	180 - 196 (179 - 199)
Integral	Likelihood = -3.03	Transmembrane	395 - 411 (392 - 412)
Integral	Likelihood = -2.55	Transmembrane	151 - 167 (151 - 168)
Integral	Likelihood = -2.02	Transmembrane	117 - 133 (117 - 133)
Integral	Likelihood = -0.64	Transmembrane	270 - 286 (269 - 286)
PERIPHERAL	Likelihood = 0.95		69

modified ALOM score: 3.30

*** Reasoning Step: 3

----- Final Results -----

Location	Confidence	Result
bacterial membrane	--- Certainty=0.6604 (Affirmative)	< succ>
bacterial outside	--- Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	--- Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

ORF01657(301 - 1596 of 1902)
 EGAD[6545]6344(1 - 434 of 439) preprotein translocase secy subunit [Lactococcus lactis]
 SP|P27148|SECY_LACLA PREPROTEIN TRANSLOCASE SECY SUBUNIT. GF|44073|emb|CAA1939.1||X59250

-2315-

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2048

A DNA sequence (GBSx2159) was identified in *S. agalactiae* <SEQ ID 6325> which encodes the amino acid sequence <SEQ ID 6326>. This protein is predicted to be 50S ribosomal protein L15 (rplO). Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

```

55      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.5259(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

60 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB54021 GB:U96620 ribosomal protein L15 [Staphylococcus aureus]

-2316-

Identities = 116/146 (79%), Positives = 128/146 (87%)

Query: 1 MKLHELKPAQSRKVRNVGRGTSSNGKTSGRGQKQKQKARSGGVRLGFEGGQTPLFRR 60
 5 Sbjct: 1 MKLHELKPAQSRKVRNVGRGTSSNGKTSGRGQKQKQKARSGGVRLGFEGGQTPLFRR 60

Query: 61 MPKRGFSNINAKYALVNLDQLNVFDGTEVTPVVLKEAGIVRAKSGVVKILNGELTKK 120
 +PKRGF+NIN KYFA+VNLDQLN FGDTEVTP +L E+G+V+ EKSG+KILANG L KK
 10 Sbjct: 61 LPKRGFTNINRKFYATVNLDQLNKFDDTEVTPFALLVBSGVVKNRKSIGIKILNGSLDKK 120

Query: 121 LSVKAARFSSKSAEAAITAKGGSIEVI 146
 L+VKA KPS SA AI AKGG+ EVI
 10 Sbjct: 121 LTVKAARFSSKSAEAAIDAKGGAHEVI 146

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6327> which encodes the amino acid sequence <SEQ ID 6328>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5329 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 135/146 (92%), Positives = 142/146 (96%)

Query: 1 MKLHELKPAQSRKVRNVGRGTSSNGKTSGRGQKQKQKARSGGVRLGFEGGQTPLFRR 60
 30 Sbjct: 1 MKLHELK ADBSRKVRNVGRGTSSNGKTSGRGQKQKQKARSGGVRLGFEGGQTPLFRR 60

Query: 61 MPKRGFSNINAKYALVNLDQLNVFDGTEVTPVVLKEAGIVRAKSGVVKILNGELTKK 120
 +PKRGF+NIN KYFALVNLDQLNVF+DGTETVP +LK+AGIVRAKSGVK+LNGELTKK
 35 Sbjct: 61 LPKRGFTNINTKFYALVNLDQLNVFDGTEVTPFALLKAGIVRAKSGVVKILNGELTKK 120

Query: 121 LSVKAARFSSKSAEAAITAKGGSIEVI 146
 L+VKAARFSSKSAEAI AKGGSIEVI
 35 Sbjct: 121 LTVKAARFSSKSAEAAITAKGGSIEVI 146

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2049

A DNA sequence (GBSx2160) was identified in *S.agalactiae* <SEQ ID 6329> which encodes the amino acid sequence <SEQ ID 6330>. Analysis of this protein sequence reveals the following:

45 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1162 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AA54020 GB:U96620 ribosomal protein L30 [Staphylococcus aureus]

Identities = 40/58 (68%), Positives = 46/58 (78%)

Query: 1 MAQIKITLTKSPIGRKPBQRKTVALGLGKLKNSVVKEDNAIRGMVAHSLVTVEE 58
 MA+++ITLTS IGR QRKTV ALGL K NSSVV EDN AIRG +N +HLVTVEE

-2317-

Subjet: 1 MAQIKITLTSRVIGRPETQRKTVEALGLKLNSSVVKEDNPAIRGQINKVHLVTVEE 58

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6331> which encodes the amino acid sequence <SEQ ID 6332>. Analysis of this protein sequence reveals the following:

5 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1088 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 56/58 (96%), Positives = 57/58 (97%)

 Query: 1 MAQIKITLTSPIGRKPEQRKTIVVVALGLKLNSSVVKEDNPAIRGMVNAISHLVTVEE 58
 MAQIKITLTSPIGRKPEQRKTIVVVALGLKLNSSVVKEDNPAIRGMV AISHLVTVE+

 Subjet: 1 MAQIKITLTSPIGRKPEQRKTIVVVALGLKLNSSVVKEDNPAIRGMVTAISHLVTVED 58

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2050

A DNA sequence (GBSx2161) was identified in *S.galactiae* <SEQ ID 6333> which encodes the amino acid sequence <SEQ ID 6334>. Analysis of this protein sequence reveals the following:

25 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3226 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2051

A DNA sequence (GBSx2162) was identified in *S.galactiae* <SEQ ID 6335> which encodes the amino acid sequence <SEQ ID 6336>. This protein is predicted to be 30S ribosomal protein S5 (rpsE). Analysis of this protein sequence reveals the following:

40 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3179 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAA22699 GB:M57621 ribosomal protein S5 [Bacillus
 stearothermophilus]

-2318-

Identities = 119/158 (75%), Positives = 139/158 (87%)

Query: 6 NAVELEERVVAINRVTKVVGGRRLRFAALVVVDGRNGRVGFGTGKAQEVPEAIRGAVEA 65
 N +ELEERVVVA+NRV KVVGGRRLRFP+AI/VVGD+NG VGFGTGKAQEVPEAIRKA+E
 5 Sbjct: 7 NKELEERVVAVNRVAKVVGGRRLRPSALVVVDKNGRVEVGFGTGKAQEVPEAIRKAIED 66

Query: 66 AKKNMVEVPMVGTTPHVEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVADITSK 125
 AKKN+EVDP+VGTTPHVEV FG +++LKPA EG GV AGS RAV+ELAG++DI SK
 10 Sbjct: 67 AKKNLIEVPTVGTTPHVEVIGFGGEBITLKPASEGTVGTAAGSPARAVIELAGISDIISK 126

Query: 126 SIGSNTPINIVRATVEGLKQLKRAEVAALRGISVSD 163
 S+GSNTPIN+VRAT+GLKQLKRAE+VA LRG +V +L
 15 Sbjct: 127 SIGSNTPINMVRATFDGLKQLKRAEDVAKLRGKTVEEL 164

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6337> which encodes the amino acid sequence <SEQ ID 6338>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3179 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 158/164 (96%), Positives = 161/164 (97%)

Query: 1 MAFKDNAVELEERVVAINRVTKVVGGRRLRFAALVVVDGRNGRVGFGTGKAQEVPEAIR 60
 MAFKDNAVELEERVVAINRVTKVVGGRRLRFAALVVVD NGRVGFOTGKAQEVPEAIR
 30 Sbjct: 1 MAFKDNAVELEERVVAINRVTKVVGGRRLRFAALVVVDNGRVGFGTGKAQEVPEAIR 60

Query: 61 KAVEAAKKNMVEVPMVGTTPHVEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVA 120
 KAVEAAKKN+EVPMVGTTPHVEV +FGGAKVLLKPAVEG+GVAAGGAVRAVIELAGVA
 35 Sbjct: 61 KAVEAAKKNMIEVPMVGTTPHVEVYINFGGAKVLLKPAVEGSGVAAGGAVRAVIELAGVA 120

Query: 121 DITSKSLGNTPINIVRATVEGLKQLKRAEVAALRGISVSDLA 164
 DITSKSLGNTPINIVRATVEGLKQLKRAEVAALRGISVSDLA
 40 Sbjct: 121 DITSKSLGNTPINIVRATVEGLKQLKRAEVAALRGISVSDLA 164

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2052

A DNA sequence (GBSx2163) was identified in *S.agalactiae* <SEQ ID 6339> which encodes the amino acid sequence <SEQ ID 6340>. This protein is predicted to be 50S ribosomal protein L18 (rplR). Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4488 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9465> which encodes amino acid sequence <SEQ ID 9466> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2319-

>GP:AA06815 GB:L47971 ribosomal protein L18 [Bacillus subtilis]
Identities = 86/120 (71%), Positives = 97/120 (80%), Gaps = 2/120 (1%)

5 Query: 4 VISKPDKNKIRQKRHRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASASTLD 63
+I+K KN R KRH RVR KLSGTA+RPRIN+FRSN IYQA+IDOV GVTLASASTLD
Sbjct: 1 MITTKTSQNAARLKIRHARVRAKLSGTAEPRINVFRSNKHIYAQIIDOVNGVTLASASTLD 60

10 Query: 64 KE--VSNSTKTEQAVVVGKLVARRAVAKGISEVVFDRGGYLYHGRVKALADAGRENGLKPF 121
K+ V+ T A VG+LVA+RA KGIS+VVFDRGGYLYHGRVKALAD+ARE GLKF
Sbjct: 61 KDLNVRSTIGUTSAATKVGKLVAKRAAKGISDVVVFDRGGYLYHGRVKALADARENGLKPF 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6341> which encodes the amino acid sequence <SEQ ID 6342>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.4488 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 116/121 (95%), Positives = 120/121 (98%)

25 Query: 1 MKIVISKPDKNKIRQKRHRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASAS 60
+KIVISKPDKNKIRQKRHRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASAS
Sbjct: 1 VKIVISKPDKNKIRQKRHRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASAS 60

30 Query: 61 TLDKEVSNSTKTEQAVVVGKLVARRAVAKGISEVVFDRGGYLYHGRVKALADAGRENGLKPF 121
TLDK+VS GTKTEQAVVVGKLVARRAVAKGISEVVFDRGGYLYHGRVKALAD+ARENGLKPF
Sbjct: 61 TLDKDVSGTKTEQAVVVGKLVARRAVAKGISEVVFDRGGYLYHGRVKALADARENGLKPF 121

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 2053

A DNA sequence (GBSx2164) was identified in *S.agalactiae* <SEQ ID 6343> which encodes the amino acid sequence <SEQ ID 6344>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1530 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA22700 GB:M57622 ribosomal protein L6 [Bacillus
stearothermophilus]
Identities = 108/178 (60%), Positives = 133/178 (74%)

50 Query: 1 MSRIQNKVITLPAGVEIINKDNVTVVKGPKGQLTRFKNWNLGIVTGEVTVTRPNDSEK 60
M R+G K I +PAGV + N VTVKPKG+LTR F+ ++ ITVEG +TVTRP+D K
Sbjct: 1 MKRVGKKPIPIAGVTVTVNGNTTVVKGPKGELTRTFPEMTITVNGNVTITVTRPDEK 60

55 Query: 61 MKTIHGTIRANLNMMVGVSGEFPKALSMRGVGYRAQLQSGKLVLVSGKSHQDEVEAPEG 120
+HGTIR+ L NMV GVS+G++KALE+ GVGYRA QG KLVLSVG SH E+E EG
Sbjct: 61 ERALHGTIRSLANVGEVSGEYKALELVGVGYRAGKQSKLVLVSGVSHPVIEPEEG 120

Query: 121 VTFVPTPTITINIVIGINKESVGQTAARVSLRSPEPTKKGIRYVGFVRPRKKGKIGK 178

-2320-

+ KVP+ T I V G +K+ VG+ AA +R++R PEPTKGGIRY GE VR KEGKTGK
 Sbjct: 121 LEIEVPSQTKIIVKGADKQVGEAANIRAVRPPEPYKGGIRYBGLVRLKBEKGTGK 178

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6345> which encodes the amino acid sequence <SEQ ID 6346>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1704 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/178 (85%), Positives = 166/178 (92%)
 Query: 1 MSRIGNKVIITLPAQVEIINKDNVITVKGPKQLTRKFNKNIGITVGEDEVIVRENDSE 60
 MSRIGNKVIITLPAQVEIINKDNVITVKGPKQLTRKFNKNIGITVGEDEVIVRENDSE 60
 Sbjct: 1 MSRIGNKVIITLPAQVELTNNNNVITVKGPKGELTRFNKNITETKVBTEITVVRPNDSE 60
 Query: 61 MKTITGTTTRANLNNMVGVBEGFKKALENRGVYRAQLQSKLVLSVGKSHQDEVEAPEG 120
 MKTITGTTTRANLNNMVGVBEGFKK LEM+GVGYRAQLQ+KLVLSVGKSHQDEVEAPEG 120
 Sbjct: 61 MKTITGTTTRANLNNMVGVBEGFKKDLNMGVYRAQLQGTGLVLSVGKSHQDEVEAPEG 120
 Query: 121 VIFEVPIPTTINIVIGINKESVQGTAAYVRSLSPEPYKGGIRYVGEVFRKEGKTGK 178
 +TF V PT+I+V GINKE VGQTAAY+RSLRSPPEPYKGGIRYVGE+VR KEGKTGK
 Sbjct: 121 ITTFVAMPITISVVGINKESVQGTAAYIRSLRSPPEPYKGGIRYVGEVFRKEGKTGK 178

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2054

A DNA sequence (GBSx2165) was identified in *S. agalactiae* <SEQ ID 6347> which encodes the amino acid sequence <SEQ ID 6348>. This protein is predicted to be 30S ribosomal protein S8 (rpsH). Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4356 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AB06813 GB:I47971 ribosomal protein S8 [Bacillus subtilis]
 Identities = 100/132 (75%), Positives = 116/132 (87%)
 Query: 1 MVMTPDIADFLTRIRANQAKHEVLEVFASNIKKGIADILKREGPVKNVEYEDDKGII 60
 MVMTPDIADFLTRIRANQAKHE LE+ES +K+ IA+ILKREGPV++VR +ED KQII
 Sbjct: 1 MVMTPDIADFLTRIRANQAKHEVLEVFASNIKKGIADILKREGPVKNVEYEDDKGII 60
 Query: 61 RVFLKYQGNRERVITLKRISKPGRLVYTKHEIMPKVINGLGIAIVSTBGLITDKKEAR 120
 RVFLKYQGN ERVIT LKRISKPGRLVY K +P+VINGLGIAI+STB+G+LTDKEAR
 Sbjct: 61 RVFLKYQGNRERVITLKRISKPGRLVYAKSNEVFRVINGLGIAIISTBQGVITDKKEAR 120
 Query: 121 KNIGGEVLAYIW 132
 K QGEVLAYIW
 Sbjct: 121 KQAGGEVLAYIW 132

-2321-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6349> which encodes the amino acid sequence <SEQ ID 6350>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4327(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 122/132 (92%), Positives = 129/132 (97%)

Query: 1  MVMIDPIADFLTRIRNANQAKHEVLEVPASNIKKGIADILKRBGPFVIONVEVIEDOKQGI 60
      15  MVMIDPIADFLTRIRNANQ KHEVLEVPASNIKKGIA+LKRBGPFVIONVEVIEDOKQGI
Sbjct: 1  MVMIDPIADFLTRIRNANQVKEVLEVPASNIKKGIATILKRBGPFVIONVEVIEDOKQGI 60

Query: 61  RVFLCYGNGERVITNLKRISKPLRVYTKHEDMPKVLKLGIAIVSTSEGLLTDKEARQ 120
      20  RVFLCYG+NGERVITNLKRISKPLRVY K +DMPKVLKLGIAI+STSEGLLTDKEARQ
Sbjct: 61  RVFLCYGNGERVITNLKRISKPLRVYAKRDEMPEVLKLGIAIISTSEGLLTDKEARQ 120

Query: 121 KNIGGEVIAYIW 132
      25  KN+GGEV+AY+W
Sbjct: 121 KNVGEVIAYVW 132

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2055

A DNA sequence (GBSx2166) was identified in *S.agalactiae* <SEQ ID 6351> which encodes the amino acid sequence <SEQ ID 6352>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CA811905 GB:Z99104 ribosomal protein S14 [Bacillus subtilis]
Identities = 47/61 (77%), Positives = 53/61 (86%)

Query: 1  MAKKSIAKNKRPAKFSQAYTRCEKGRPHSVYRKPFQLCRCVRLDAYKGQVPGVTKAS 60
      45  MAKKSIAK +R KF Q YTRCE+GRPHSV RKF+LCR+CFR+LAYKGQ+PGV KAS
Sbjct: 1  MAKKSIAKQQRTPKFKVQYTRCERCGRPHSVIRKFKLCRICFRLAYKGQIPGVTKAS 60

Query: 61 W 61
      W
Sbjct: 61 W 61

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6353> which encodes the amino acid sequence <SEQ ID 6354>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

-2322-

bacterial cytoplasm --- Certainty=0.4747 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 55/61 (90%), Positives = 59/61 (96%)

Query: 1 NAKSMIAKNKRPKAKFSTQAYTRCEKOGPHSVYRKFLCRVCFRDLAYKQQVPGVTKAS 60

10 Sbjct: 1 LAKSMIAKNKRPKAKHSTQAYTRCEKOGPHSVYRKFLCRVCFRELAYKQIIPGVTKAS 60

Query: 61 W 61

W

Sbjct: 61 W 61

15

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2056

20 A DNA sequence (GBSx2167) was identified in *S. agalactiae* <SEQ ID 6355> which encodes the amino acid sequence <SEQ ID 6356>. This protein is predicted to be 50S ribosomal protein L5 (rplE). Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1845 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB03865 GB:AP001507 ribosomal protein L5 (BL6) [Bacillus halodurans]
 Identities = 143/178 (80%), Positives = 162/178 (90%)

Query: 3 NRLKEKYTNVEVPALTEKFMYSVMVAPKVKIVLNNVGDAVSNNAKLEKAAAEALIS 62

35 NRLKEKY E+VP+LTERFMYSVMVAPK+EKIV+NNVGDAV NAK L+KA EL I+

Sbjct: 2 NRLKEKYQKELVPSLTERFMYSVMVAPKLEKIVNNVGDAVQNAKALDKAVEELTIT 61

Query: 63 GQKPLITAKKSIAGFLRBCVAIGAKYTLRGERMYEFLDKLVSVSLPRVDFHGVPTKS 122

40 GQKPLITAKKSIAGFLRBCVAIGAKYTLRGERMYEFLDKLVSVSLPRVDFHGVPTKS 122

Sbjct: 62 GQKPLITAKKSIAGFLRBCVAIGAKYTLRGERMYEFLDKLVSVSLPRVDFHGVPTKS 121

Query: 123 FDGRGNVTLGVKEQLIFPEINFDVDKVRGLDIVVTANTDEERRELLKGLMPFAK 180

FDGRGNVTLGVKEQLIFPEI+D VDKVRG+D+VIVTTR+TDEE+RELL +GMPP K

Sbjct: 122 FDGRGNVTLGVKEQLIFPEIDYDKVDKVRGMDDVIVTTR+TDEE+RELLSQGMHPPQR 179

45

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6357> which encodes the amino acid sequence <SEQ ID 6358>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1793 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

Identities = 177/180 (98%), Positives = 180/180 (99%)

-2323-

Query: 1 MANRLKEKYTNVVPALTEKPNYSVMVAVKVEKIVLNMGVGDVSNKLNLEKAAAEAL 60
 MANRLKEKYTNVVPALTEKPNYSVMVAVKVEKIVLNMGVGDVSNKLNLEKAAAEAL
 Sbjet: 1 MANRLKEKYTNVVPALTEKPNYSVMVAVKVEKIVLNMGVGDVSNKLNLEKAAAEAL 60

Query: 61 ISGQKPLITAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVP 120
 ISGQKPLITAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVP
 Sbjet: 61 ISGQKPLITAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSVSLPRVRDFHGVP 120

Query: 121 KSFDGRGNYTLGVKEQLIFFEIPFDVDRVGLDIVIVTTANTDESKRLKGLGMPFAK 180
 KSFDGRGNYTLGVKEQLIFFEIPFDVDRVGLDIVIVTTANTDESKRLKGLGMPFAK
 Sbjet: 121 KSFDGRGNYTLGVKEQLIFFEIPFDVDRVGLDIVIVTTANTDESKRLKGLGMPFAK 180

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2057

A DNA sequence (GBSx2169) was identified in *S. agalactiae* <SEQ ID 6359> which encodes the amino acid sequence <SEQ ID 6360>. This protein is predicted to be 50S ribosomal protein L24 (rplX). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1850 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA033285 GB:AF126061 Rpl24 [Streptococcus pneumoniae]
 Identities = 89/101 (88%), Positives = 94/101 (92%)

Query: 1 MFVKKGDKVRVLAGKDKGTEAVVLKALPKVKNKVVSGVALIKKHQKPNENPQGAIVEKE 60
 MFVKKGDKVRVLAGKDKGTEAVVL ALPKVKNKVVSGV +KKHQ+P NE PQG I+EKE
 Sbjet: 1 MFVKKGDKVRVLAGKDKGTEAVVLALPKVKNKVVSGVNIKKHQRTNELPQGGIIEKE 60

Query: 61 APIHVSIVQVLKNGVAGRVGYKFDGKKVRYNKSSEVL 101
 A IHVSIVQVLKNGVAGRVGYK VDGKKVRYNKSSEVL
 Sbjet: 61 APIHVSIVQVLKNGVAGRVGYKFDGKKVRYNKSSEVL 101

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6361> which encodes the amino acid sequence <SEQ ID 6362>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1850 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 95/101 (94%), Positives = 99/101 (97%)

Query: 1 MFVKKGDKVRVLAGKDKGTEAVVLKALPKVKNKVVSGVALIKKHQKPNENPQGAIVEKE 60
 MFVKKGDKVRVLAGKDKGTEAVVLKALPKVKNKVVSGV +KKHQKPNENPQGAIVEKE
 Sbjet: 1 MFVKKGDKVRVLAGKDKGTEAVVLKALPKVKNKVVSGVGMKKHQKPNENPQGAIVEKE 60

Query: 61 APIHVSIVQVLKNGVAGRVGYKFDGKKVRYNKSSEVL 101
 APIHVSIVQVLKNGVAGRVGYKFDGKKVRYNKSSEVL

-2324-

Subjct: 61 APIHVSNVQVLDKNGVAGRIGYKVVDDGKKVYKSKSGEVLDD 101

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2058

A DNA sequence (GBSx2170) was identified in *S. agalactiae* <SEQ ID 6363> which encodes the amino acid sequence <SEQ ID 6364>. This protein is predicted to be 50S ribosomal protein L14 (rplN). Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1004 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA033284 GB:AF126061 RplL14 [Streptococcus pneumoniae]
 Identities = 116/122 (95%), Positives = 120/122 (98%)

20 Query: 1 MIQCBTELKIVADNSGAREILTIKVLGSGGRKFANIGDVIASVKGQATPGGAVKKGDDVKA 60
 MIQCBTELKIVADNSGAREILTIKVLGSGGRKFANIGDVIASVKGQATPGGAVKKGDDVKA
 Subjct: 1 MIQCBTELKIVADNSGAREILTIKVLGSGGRKFANIGDVIASVKGQATPGGAVKKGDDVKA 60

25 Query: 61 VIVRTKIGARRPDGSGYIKFDNNAAVIIRDDKTPRGTRIPGPVARELBSGGYMKIVSLAPE 120
 VIVRTKIGARRPDGSGYIKFDNNAAVIIRDDKTPRGTRIPGPVARELBSGGYMKIVSLAPE
 Subjct: 61 VIVRTKIGARRPDGSGYIKFDNNAAVIIRDDKTPRGTRIPGPVARELBSGGYMKIVSLAPE 120

30 Query: 121 VL 122
 VL 122
 Subjct: 121 VL 122

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6365> which encodes the amino acid sequence <SEQ ID 6366>. Analysis of this protein sequence reveals the following:

35 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1004 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 122/122 (100%), Positives = 122/122 (100%)

45 Query: 1 MIQCBTELKIVADNSGAREILTIKVLGSGGRKFANIGDVIASVKGQATPGGAVKKGDDVKA 60
 MIQCBTELKIVADNSGAREILTIKVLGSGGRKFANIGDVIASVKGQATPGGAVKKGDDVKA
 Subjct: 1 MIQCBTELKIVADNSGAREILTIKVLGSGGRKFANIGDVIASVKGQATPGGAVKKGDDVKA 60

50 Query: 61 VIVRTKIGARRPDGSGYIKFDNNAAVIIRDDKTPRGTRIPGPVARELBSGGYMKIVSLAPE 120
 VIVRTKIGARRPDGSGYIKFDNNAAVIIRDDKTPRGTRIPGPVARELBSGGYMKIVSLAPE
 Subjct: 61 VIVRTKIGARRPDGSGYIKFDNNAAVIIRDDKTPRGTRIPGPVARELBSGGYMKIVSLAPE 120

55 Query: 121 VL 122
 VL 122
 Subjct: 121 VL 122

-2325-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2059

A DNA sequence (GBSx2171) was identified in *S. agalactiae* <SEQ ID 6367> which encodes the amino acid sequence <SEQ ID 6368>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3415 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD33283 GB: AF126061 RpS17 [Streptococcus pneumoniae]
Identities = 82/86 (95%), Positives = 83/86 (96%)

Query: 1 MERNQKTLYGKRVVSDKMDKTIITVVETKRNHPVYIGKRLINYSKKYKAHDENNVAKEGDIV 60
MERN RK L GRVUSDKMDKTIITVVETKRNHPVYIGKRLINYSKKYKAHDENNVAKEGDIV
Sbjct: 1 MERNNRKVLVGRVUSDKMDKTIITVVETKRNHPVYIGKRLINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRFLSATKRFRFLVEVVEKAVII 86
RIMETRFLSATKRFRFLVEVVEKAVII
Sbjct: 61 RIMETRFLSATKRFRFLVEVVEKAVII 86

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6369> which encodes the amino acid sequence <SEQ ID 6370>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3415 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/86 (100%), Positives = 86/86 (100%)

Query: 1 MERNQKTLYGKRVVSDKMDKTIITVVETKRNHPVYIGKRLINYSKKYKAHDENNVAKEGDIV 60
MERNQKTLYGKRVVSDKMDKTIITVVETKRNHPVYIGKRLINYSKKYKAHDENNVAKEGDIV
Sbjct: 1 MERNQKTLYGKRVVSDKMDKTIITVVETKRNHPVYIGKRLINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRFLSATKRFRFLVEVVEKAVII 86
RIMETRFLSATKRFRFLVEVVEKAVII
Sbjct: 61 RIMETRFLSATKRFRFLVEVVEKAVII 86

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2060

A DNA sequence (GBSx2172) was identified in *S. agalactiae* <SEQ ID 6371> which encodes the amino acid sequence <SEQ ID 6372>. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

-2326-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4329 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33282 GB:AF126061 RpL29 [Streptococcus pneumoniae]

Identities = 58/68 (85%), Positives = 64/68 (93%)

Query: 1 MKLQELKDFVKELRGLSQBELAKENELAKKELFDLRFQAAOQLEKTARLDEVKKQTARV 60
 MKL E+K+PVKELRGLSQBELAK+ENELAKKELF+LRFQAA OQLE+TARL EVKKQTAR+
 Sbjct: 1 MKLNEVKSPVKELRGLSQBELAKPENLAKKELFELRFPQAATQQLKQTARLCKVKQTARI 60

Query: 61 KTVQSEMK 68

KTVQSE K

Sbjct: 61 KTVQSEK 68

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2061

A DNA sequence (GBSx2174) was identified in *S. agalactiae* <SEQ ID 6373> which encodes the amino acid sequence <SEQ ID 6374>. This protein is predicted to be RpL16 (rplP). Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4574 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33263 GB:AF126059 RpL16 [Streptococcus pneumoniae]

Identities = 135/137 (98%), Positives = 137/137 (99%)

Query: 1 MLVFRVRRHRRFRGKMRGKAGKGEVSPGEYGLQATTSHWITNQIEAARIAMTRYHGR 60
 MLVFRVRRHRRFRGKMRGKAGKGEV+PGEYGLQATTSHWITNQIEAARIAMTRYHGR
 Sbjct: 1 MLVFRVRRHRRFRGKMRGKAGKGEVAPGEYGLQATTSHWITNQIEAARIAMTRYHGR 60

Query: 61 GKGWIKIFPHKSYTAKAIGVRMSSGKGAFBGWVAPVKGKVMFBIAGVSEEVAREALRL 120
 GKGWIKIFPHKSYTAKAIGVRMSSGKGAFBGWVAPVKGKVMFBIAGVSE+AREALRL
 Sbjct: 61 GKGWIKIFPHKSYTAKAIGVRMSSGKGAFBGWVAPVKGKVMFBIAGVSEELAREALRL 120

Query: 121 ASHKLPVKCKFVKREAR 137

ASHKLPVKCKFVKREAR

Sbjct: 121 ASHKLPVKCKFVKREAR 137

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6375> which encodes the amino acid sequence <SEQ ID 6376>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4574 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2327-

Identities = 136/137 (99%), Positives = 137/137 (99%)

Query: 1 MLVPKRVKRRREFRGKMGKRGKGGKEVSFGVEYGLQATTSHWITNRQIEAARIAMTRYMKR 60
 5 Sbjct: 1 MLVPKRVKRRREFRGKMGKRGKGGKEVSFGVEYGLQATTSHWITNRQIEAARIAMTRYMKR 60

Query: 61 GGGKWKIKIPHKSYTAKAIGVRMGSGKGAPBGWVAPVKRGKVMFEIAGVSEEVAREALRL 120
 10 Sbjct: 61 GGGKWKIKIPHKSYTAKAIGVRMGSGKGAPBGWVAPVKRGKVMFEIAGVSEEVAREALRL 120

Query: 121 ASHKLPVKCKFVKREAR 137
 ASHKLPVKCKFVKREAR
 10 Sbjct: 121 ASHKLPVKCKFVKREAR 137

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2062

A DNA sequence (GBSx2175) was identified in *S. agalactiae* <SEQ ID 6377> which encodes the amino acid sequence <SEQ ID 6378>. Analysis of this protein sequence reveals the following:

20 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.3756 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAD33280 GB:AP126061 Rp63 [Streptococcus pneumoniae]
 Identities = 200/208 (96%), Positives = 203/208 (97%)

Query: 10 MRVGIIRDWDAKQYAEKEYADYLHEDLAIRKPKELADA+VSTIETIERAVNKVIVSLHT 69
 35 Sbjct: 1 MRVGIIRDWDAKQYAEKEYADYLHEDLAIRKPKELADA+VSTIETIERAVNKVIVSLHT 60

Query: 70 AKPGMVGIGGAGNVDAIRGLQNLKLTGKQVHINIIEIKQPDLDALHVGENTIAQLQEQRVAF 129
 35 Sbjct: 61 AKPGMVGIGGAGNVDAIRGLQNLKLTGKQVHINIIEIKQPDLDALHVGENTIAQLQEQRVAF 120

Query: 130 RRAQKQAIQRTRMRAGAKGIKTQVSGRLNGADIAAREGYSEGTVPLHTLRADIDYANEAD 189
 40 Sbjct: 121 RRAQKQAIQRTRMRAGAKGIKTQVSGRLNGADIAAREGYSEGTVPLHTLRADIDYANEAD 180

Query: 190 TTYGKLGKVKWIYRGSEVLEARKNTGKGK 217
 45 TTYGKLGKVKWIYRGSEVLEARKNTGKGK
 Sbjct: 181 TTYGKLGKVKWIYRGSEVLEARKNTGKGK 208

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6379> which encodes the amino acid sequence <SEQ ID 6380>. Analysis of this protein sequence reveals the following:

50 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.3756 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2328-

Example 2063

A DNA sequence (GBSx2176) was identified in *S.agalactiae* <SEQ ID 6381> which encodes the amino acid sequence <SEQ ID 6382>. This protein is predicted to be 50S ribosomal protein L22 (rplV). Analysis of this protein sequence reveals the following:

```

5   Possible site: 36
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2704 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:AAD33279 GB:AF126061 Rpl22 [Streptococcus pneumoniae]
    Identities = 99/114 (86%), Positives = 106/114 (92%)

    Query: 1  MAEITSAKAMARTVRSVKTRLVLDLRGKQVADAIAILKFTPNKAARVIEKTLNSAIA 60
              MAEITSAKAMARTVRSVKTRLVLD LRKQVADAIAIL FTPNKA A +I K LNSA+A
    Sbjct: 1  MAEITSAKAMARTVRSVKTRLVLDLRGKSVADAIAILFTPNKAAEILKVLNSAVA 60

20   Query: 61 NAENNFGLEKANLAVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
              NAENNFGLEKANLAVVSE FANEGPTMKRFRPRAKGSASPINKRT H+TV V+EK
    Sbjct: 61 NAENNFGLEKANLAVVSEAFANEGPTMKRFRPRAKGSASPINKRTAHITVAVAEK 114

```

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6383> which encodes the amino acid sequence <SEQ ID 6384>. Analysis of this protein sequence reveals the following:

```

    Possible site: 36
    >>> Seems to have no N-terminal signal sequence

30   ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.2794 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

35 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 113/114 (99%), Positives = 113/114 (99%)

    Query: 1  MAEITSAKAMARTVRSVKTRLVLDLRGKQVADAIAILKFTPNKAARVIEKTLNSAIA 60
              MAEITSAKAMARTVRSVKTRLVLDLRGK VADAIAILKFTPNKAARVIEKTLNSAIA
40   Sbjct: 1  MAEITSAKAMARTVRSVKTRLVLDLRGKQVADAIAILKFTPNKAARVIEKTLNSAIA 60

    Query: 61 NAENNFGLEKANLAVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114
              NAENNFGLEKANLAVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK
    Sbjct: 61 NAENNFGLEKANLAVVSETFANEGPTMKRFRPRAKGSASPINKRTTHVTVVVSEK 114

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2064

50 A DNA sequence (GBSx2177) was identified in *S.agalactiae* <SEQ ID 6385> which encodes the amino acid sequence <SEQ ID 6386>. This protein is predicted to be 30S ribosomal protein S19 (rpsS). Analysis of this protein sequence reveals the following:

```

    Possible site: 23
    >>> Seems to have no N-terminal signal sequence

55   ----- Final Results -----

```

-2329-

bacterial cytoplasm --- Certainty=0.2991 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein is similar to ribosomal protein S19 from *S.pneumoniae*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6387> which encodes the amino acid sequence <SEQ ID 6388>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3319 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 92/92 (100%), Positives = 92/92 (100%)

Query: 1 MGRSLKKGPFVDEHLMKKVEAQANDEKKVVKTWSSRRSTIFPSFIGYITIAVYDGRKHVPV 60
 MGRSLKKGPFVDEHLMKKVEAQANDEKKVVKTWSSRRSTIFPSFIGYITIAVYDGRKHVPV
 Sbjct: 19 MGRSLKKGPFVDEHLMKKVEAQANDEKKVVKTWSSRRSTIFPSFIGYITIAVYDGRKHVPV 78

Query: 61 YIQEDMVGHKLGEFAPTRTYKHAADKKTRR 92
 YIQEDMVGHKLGEFAPTRTYKHAADKKTRR
 Sbjct: 79 YIQEDMVGHKLGEFAPTRTYKHAADKKTRR 110

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2065

30 A DNA sequence (GBSx2178) was identified in *S.agalactiae* <SEQ ID 6389> which encodes the amino acid sequence <SEQ ID 6390>. This protein is predicted to be L2 (rplB). Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3182 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA45959 GB:U43929 L2 [Bacillus subtilis]
 Identities = 208/277 (75%), Positives = 239/277 (86%)

45 Query: 1 MGKIKYKPTNCRNMSTSLDFAITTTPEKSLNLSKNKAGRWNRITVVRHQGGHKK 60
 M IK YKP+NGRR MT+ DFAITTT PEKSL L K GRW G++TVVRHQGGHKK
 Sbjct: 1 MAIKKYPKPSNGRRGMTTSDFAITTTPEKSLAPLHKKGGRNQGKLTVVRHQGGHKK 60

50 Query: 61 HYRLIDFKRNKDGVEAVVKTTSYDPNRTANIALVHYITGVKAYILAPKGLFVQRTIISGP 120
 YR+IDFKR+KDG+ V T+EYDPR+ANIAL+Y DG K YILAPKGL+VQ ++SGP
 Sbjct: 61 QYRVDFCRKDKGIPGRVATVSYDPNRSANIALINLYADGEKRYTILAPKGLVQGTSTVMSGP 120

Query: 121 EADIKVGNALPLNIPFGTVVHNTELQPGKGAELIRAGASAQVLQGRKYLVLRLQSGE 180
 EADIKVGNALPL NIPFGTV+HNTEL+PGK +L+R+AG SAQVLG+EKYVLVLRL QGS
 55 Sbjct: 121 EADIKVGNALPLNIPFGTVVHNTELKPKGGGLVRSAGTSAQVLGEEKYVLVLRLQSGE 180

Query: 181 VRMILGTCRATIGTVGNBQGLVNIWIKAGRNKRGVVRPTVGRGWNPNDFHGGGKAP 240

-2330-

VRNII CRA+IG VGNRQ I+NIHGKAGR+RWKG+RPTVRGSMVNPNDHPHGGGEG+AP
 Sbjct: 181 VRNIIACRASIGVGNRQHELINIGKAGSRWKGIRPTVRGSMVNPNDHPHGGGEGRAP 240

Query: 241 VGRKAPSTPMGKPALGLKTRNKKAKSKLIVRRNRK 277
 +GRK+P +PMGK LG KTR KK KSKD IVRNR K
 Sbjct: 241 IGRKSPMSPMGKPTLGFKTRKGGNKSDFIVRRNRK 277

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6391> which encodes the amino acid sequence <SEQ ID 6392>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2560 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 264/277 (95%), Positives = 276/277 (99%)

Query: 1 MGIIKVKYPTTNGRFRNMTSLDFABITTTPEKSLVSLKNGKAGNNNGRITVRHQGGGHR 60
 +GIKVKYPTTNGRFRNMTSLDFABITTTPEKSLVSLK+KAGNNNGRITVRHQGGGHR
 Sbjct: 1 VGIIKVKYPTTNGRFRNMTSLDFABITTTPEKSLVSLKNGKAGNNNGRITVRHQGGGHR 60

Query: 61 HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYILAPKGLVGGRIISGP 120
 HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYI+APKGLVGGRI+SGP
 Sbjct: 61 HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYIIAPKGLVGGRI+VSGP 120

Query: 121 EADIKVGNALFLANIPVGTVHNIELQPKGAGELIARAASACVLQGGGKYLVRLQSGE 180
 +ADIKVGNALFLANIPVGTV+HNIEL+PKG EL+ARAASACVLQGGGKYLVRLQSGE
 Sbjct: 121 DADIKVGNALFLANIPVGTVHNIELQPKGAGELIARAASACVLQGGGKYLVRLQSGE 180

Query: 181 VRNIIIGTCRATIGTVGNEQQSLVNIGKAGRNKRGVRRPTVRGSMVNPNDHPHGGGEGKAP 240
 VRNIIIGTCRATIGTVGNEQQSLVNIGKAGR+RWKG+RPTVRGSMVNPNDHPHGGGEGKAP
 Sbjct: 181 VRNIIIGTCRATIGTVGNEQQSLVNIGKAGSRWKGIRPTVRGSMVNPNDHPHGGGEGKAP 240

Query: 241 VGRKAPSTPMGKPALGLKTRNKKAKSKLIVRRNRK 277
 VGRKAPSTPMGKPALGLKTRNKKAKSKLIVRRNR+K
 Sbjct: 241 VGRKAPSTPMGKPALGLKTRNKKAKSKLIVRRNRK 277

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2066

A DNA sequence (GBSx2180) was identified in *S.agalactiae* <SEQ ID 6393> which encodes the amino acid sequence <SEQ ID 6394>. This protein is predicted to be 50S ribosomal protein L23 (rp1W). Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1669 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA03855 GB:AP001507 ribosomal protein L23 [Bacillus halodurans]
 Identities = 56/92 (60%), Positives = 67/92 (71%), Gaps = 1/92 (1%)

-2331-

Query: 2 NLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLILKQAVEAARDGVKVASVNTVTVK 61
 N DVIK+EVITE+S + KYTPEVD RA+K IK A+E ED VKVA VNT+ K
 Sbjct: 3 NARDVIKRFPVITERSTSWGDKKYTFEVDVRANKTQIKDAIEEIPD-VKVAKNTMNYG 61

5 Query: 62 KAKRVGRYTGFTSKTKKAIITLTADSKAIELF 93
 K KR GRYTGT++ KKAI+TLT DSK ++ F
 Sbjct: 62 KPRKRGYTGFTARRKKAITVLTADSKELDF 93

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6395> which encodes the amino acid
 10 sequence <SEQ ID 6396>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1617 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 96/98 (97%), Positives = 97/98 (98%)

Query: 1 NMLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLILKQAVEAARDGVKVASVNTVTVK 60
 NMLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLILKQAVEAARDGVKVASVNTV VK
 25 Sbjct: 1 NMLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLILKQAVEAARDGVKVASVNTVVK 60

Query: 61 KAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98
 KAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE
 Sbjct: 61 KAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 2067

A DNA sequence (GBSx2181) was identified in *S.agalactiae* <SEQ ID 6397> which encodes the amino
 acid sequence <SEQ ID 6398>. This protein is predicted to be 50S ribosomal protein L4 (rplD). Analysis of
 35 this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.54 Transmembrane 140 - 156 (139 - 156)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.1617 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45957 GB:U43929 L4 [Bacillus subtilis]
 Identities = 130/207 (62%), Positives = 160/207 (76%)

50 Query: 1 MANVGLFDQGTGKEVSSVLENEAIPGIEPNESVVFDFVIGQRASLRQOTHAVKNRSVSGG 60
 M V L+Q G +ELN ++PGIEPNESVVPD ++ QRASLRQOTH VKNRS V GG
 Sbjct: 1 MPKVALYNGCSTAGDIELNASVFGISPNESVVFDAIIMQRASLRQOTHKVRNRSVRRG 60

Query: 61 GRKFWRKQGTGRARQGSIRSPQWRGGGVGFTPRSYGYLKPQKVRRLALKSVYSKVAE 120
 GRKFWRKQGTGRARQGSIRSPQWRGGGVGFTPRSY YKLP+KVRRLA+KSV S+KV +
 55 Sbjct: 61 GRKFWRKQGTGRARQGSIRSPQWRGGGVGFTPRSYGYLKPQKVRRLALKSVLSSKVYD 120

Query: 121 DKPVAVENLGFAPKPAEFAFVLSALGIDSKVLVILEEGNEFAALGARLEPNVTATATT 180
 + ++E+L+ KT K A++L LS++ K L++ + NE ALSARN+P VTV A

-2332-

Subjct: 121 INIIIVLEDTLDTAKTKKMAILLKGLSVEKKALIVTADANEAVALSARNIPGVTVVEANG 180

Query: 181 ASVLDIVNADKLLVTKKAISTIEGVLA 207
+VLD+VN +KLL+TK A+ +E VLA

Subjct: 181 INVLDIVNHEKLLITKAAVKEVLEVA 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6399> which encodes the amino acid sequence <SEQ ID 6400>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2544 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 199/207 (96%), Positives = 203/207 (97%)

Query: 1 MANVKLFEDQTGKEVSSVELNEAIPGISPNESVVPDVISQRAIRQGTTHAVKNRSVAVSGG 60

Subjct: 1 MANVKLFEDQTGKEVSSVELNDAIPGISPNESVVPDVISQRAIRQGTTHAVKNRSVAVSGG 60

Query: 61 GRKFWRKQGTGRARQGSIRSPQWRGGGVVFGPTPRSYGKLPQKVRRLALKSVYSKVAE 120

Subjct: 61 GRKFWRKQGTGRARQGSIRSPQWRGGGVVFGPTPRSYGKLPQKVRRLALKSVYSKVAE 120

Query: 121 DKPVAVENLSFAAPKTAEFASVLSALSDSKVIVLVEEGNEFAALSARNLPNVTVATATT 180

Subjct: 121 DKPVAVEGLSFAAPKTAEFKVLALSISDITKVIDLVEEGNEFAALSARNLPNVTVATATT 180

Query: 181 ASVLDIVNADKLLVTKKAISTIEGVLA 207

Subjct: 181 ASVLDIVNADKLLVTKKAISTIEGVLA 207

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2068

A DNA sequence (GBSx2183) was identified in *S.agalactiae* <SEQ ID 6401> which encodes the amino acid sequence <SEQ ID 6402>. This protein is predicted to be 50S ribosomal protein L3 (rplC). Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2090 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45956 GB:U43929 L3 [Bacillus subtilis]

Identities = 157/208 (75%), Positives = 180/208 (86%), Gaps = 2/208 (0%)

Query: 1 MTKGILGKKVGMITQIPTESGEPIPVTVIERTPNVVLQVTVETDGYEAVQVGFDDKREVL 60

Subjct: 1 MTKGILGK+K+GMTQ+P E+G+ IPIVTVIER PNVVLQ KT E DGYE+Q+GFDKRE L

Subjct: 1 MTKGILGRKIGMTQVPAENGDLIPVTVIERAPNVVLQKTAENDGYEAIQLGFDKREKL 60

Query: 61 SNKPAKHVAKANTAPKRFIRKFIQIE--GLEVGAEISVEQFAGDVVDVTGTSKGGQPO 118

-2333-

SNKP KGHVAKA TAPKRP++E + +E EVG E+ VE F AG++VDVIG SKGKGFG
 5 Sbjct: 61 SNKPEKGHVAKAETAPKRPVKELRGVEMDAYEVGQEVKVEIFSAEIVDVTVGSVSKGKGFG 120

Query: 119 GVIKRHQSGRGPMAGHSRYHRRPGSMGPVAPNRVFNKRLAGRMGNGNRVTVNLEIVQVI 178
 G IKRHQSGRGPM+HGSRYHRRPGSMGPV PNRVFK K L GRMGG ++TVQNLEIV+V
 Sbjct: 121 GAIKRGHSGRGPM+HGSRYHRRPGSMGPVDPNIRVFGKLLPGRMGGBQITVQNLEIVKVD 180

Query: 179 PEKNVVLKGNVPGAKSLITIKSAVKA 206
 E+H++LIKGNVPGAKSLIT+KSAVK+
 10 Sbjct: 181 AERNLLIKGNVPGAKSLITVKSARKS 208

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6403> which encodes the amino acid sequence <SEQ ID 6404>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2090 (Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 205/208 (98%), Positives = 207/208 (98%)

25 Query: 1 MTKGILLKKGVMGTQIFTSGEFIPVTVEATPNNVLQVKTVETDGYEAVQVGFDDKREVL 60
 MTKGILLKKGVMGTQIFTSGEFIPVTVEATPNNVLQVKTVETDGYEAVQVGFDDKREVL
 Sbjct: 1 MTKGILLKKGVMGTQIFTSGEFIPVTVEATPNNVLQVKTVETDGYEAVQVGFDDKREVL 60

30 Query: 61 SNKPAKSHVAKANTAPKRFIREFKNIEGLEVGAELSVQFEAGDVVDVTGSKGKGFGV 120
 SNKPAKSHVAKANTAPKRFIREFKNIEGLEVGAELSVQFEAGDVVDVTG SKGKGFGV
 Sbjct: 61 SNKPAKSHVAKANTAPKRFIREFKNIEGLEVGAELSVQFEAGDVVDVTGSKGKGFGV 120

Query: 121 IKRHQSGRGPMAGHSRYHRRPGSMGPVAPNRVFNKRLAGRMGNGNRVTVNLEIVQVI 180
 IKRHQSGRGPMAGHSRYHRRPGSMGPVAPNRVFNKRLAGRMGNGNRVTVNLEIVQVI
 35 Sbjct: 121 IKRHQSGRGPMAGHSRYHRRPGSMGPVAPNRVFNKRLAGRMGNGNRVTVNLEIVQVI 180

Query: 181 KNNVLKGNVPGAKSLITIKSAVKA 208
 KNN+L+KGNVPGAKSLITIKSAVKA
 40 Sbjct: 181 KNNVLKGNVPGAKSLITIKSAVKA 208

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2069

A DNA sequence (GBSx2184) was identified in *S.agalactiae* <SEQ ID 6405> which encodes the amino acid sequence <SEQ ID 6406>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.43 Transmembrane 5 - 21 (5 - 21)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1171 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2334-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2070

A DNA sequence (GBSx2185) was identified in *S.agalactiae* <SEQ ID 6407> which encodes the amino acid sequence <SEQ ID 6408>. This protein is predicted to be 30S ribosomal protein S10 (rpsJ). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3160 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA846363 GB:L29637 S10 ribosomal protein [Streptococcus mutans]
Identities = 98/102 (96%), Positives = 102/102 (99%)

Query: 1 MANKKIRIRLKAYEHRITLDTAAEKIVETATRTGATVAGVPVLPTERSLEYTIIRATHKYKD 60
MANKKIRIRLKAYEHRITLDTAAEKIVETATRTGA-VAGVPVLPTERSLEYTIIRATHKYKD
Sbjct: 1 MANKKIRIRLKAYEHRITLDTAAEKIVETATRTGASVAGVPVLPTERSLEYTIIRATHKYKD 60

Query: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPGGVNVVEIKL 102
SREQFEMRTHKRL+DI+NPQTQKTVDALMKLDLPGGVNVVEIKL
Sbjct: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPGGVNVVEIKL 102

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6409> which encodes the amino acid sequence <SEQ ID 6410>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3160 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 102/102 (100%), Positives = 102/102 (100%)

Query: 1 MANKKIRIRLKAYEHRITLDTAAEKIVETATRTGATVAGVPVLPTERSLEYTIIRATHKYKD 60
MANKKIRIRLKAYEHRITLDTAAEKIVETATRTGATVAGVPVLPTERSLEYTIIRATHKYKD
Sbjct: 1 MANKKIRIRLKAYEHRITLDTAAEKIVETATRTGATVAGVPVLPTERSLEYTIIRATHKYKD 60

Query: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPGGVNVVEIKL 102
SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPGGVNVVEIKL
Sbjct: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPGGVNVVEIKL 102

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2071

A DNA sequence (GBSx2186) was identified in *S.agalactiae* <SEQ ID 6411> which encodes the amino acid sequence <SEQ ID 6412>. Analysis of this protein sequence reveals the following:

Possible site: 34

-2335-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2538 (Affixative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2072

A DNA sequence (GBSx2187) was identified in *S.agalactiae* <SEQ ID 6413> which encodes the amino acid sequence <SEQ ID 6414>. Analysis of this protein sequence reveals the following:

15 Possible site: 46
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.41	Transmembrane	88 - 104 (79 - 110)
INTEGRAL	Likelihood = -8.39	Transmembrane	304 - 320 (300 - 324)
INTEGRAL	Likelihood = -6.58	Transmembrane	185 - 201 (180 - 206)
INTEGRAL	Likelihood = -5.63	Transmembrane	338 - 354 (331 - 357)
INTEGRAL	Likelihood = -5.52	Transmembrane	240 - 256 (237 - 259)
INTEGRAL	Likelihood = -4.99	Transmembrane	383 - 399 (375 - 407)
INTEGRAL	Likelihood = -3.82	Transmembrane	49 - 65 (48 - 73)
INTEGRAL	Likelihood = -2.87	Transmembrane	127 - 143 (121 - 144)
INTEGRAL	Likelihood = -2.81	Transmembrane	159 - 175 (159 - 177)
INTEGRAL	Likelihood = -2.18	Transmembrane	30 - 46 (30 - 47)

20 ----- Final Results -----

30 bacterial membrane --- Certainty=0.5564 (Affixative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:BA806655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 132/423 (31%), Positives = 210/423 (49%), gaps = 16/423 (3%)

Query: 7 IYLAIPAMIEMLQMLMGVVVDNYLVAOLGVAVSGVSVANNIITYQAIF--IALGASI 64
 L P IE +L M M G D ++Q AV+ V V+N I+ + +F +A G SI
 Sbjct: 11 LFAITWPIPIETILMLMKNADTLMISQYSDDAVAVGVSNQILAVITVMPGFVATGSI 70

40 Query: 65 ASLLAKSLAGSKDDAISVCSQAFITLLIGAVLSIISIVFGOTFFKLLGTTKSVQVGG 124
 L+A+ L +++++A V +I L+ G VLG+ I FG K+ S+ Q
 Sbjct: 71 --LVAQLGAKERINAGKAVVVSIGANLPIGVILGLLLIAPGPPIKAMQLDLSLQSAT 126

45 Query: 125 LYLALVGGGVVTLGNLITLSSFLRVQGPRLMYVSVFNFLNAVLSGFAIFEW---Y 180
 LYL IVGG V ++ T G+ LR + MYV+I +N LN + + IF
 Sbjct: 129 LYLQVVGGSVVQSLINTAGAILRSHSPTKDVMTYTGNNILNIGNLFLIPGPGFIPVL 188

50 Query: 181 GLGVAVSTLIARLIGICILAKYL-----PIKKIKRMWKISAOIWNLALPSAGER 232
 G+ GVA+ST+++R IG+ +A L P ++KR + + +PSAGE+
 Sbjct: 189 GVTGVALSTVSTRTGLSVIALLIKRIRGLGPFAYLLKGFPRVERNNLLKIGIPSGAGD 248

Query: 233 LNMAGDVVIVAIIVVOLGTNVVAGNAIGETLTQFNFMGLGIATATITLAKYVGQKNRE 292
 L A +VI + +GT+ + L F ++ + I T L VG K +
 Sbjct: 249 LSYNAQLVITYPIAMGTALATKVTQNLMMFVLFAVALGQGTLLIGHQVAGKIQ 308

55 Query: 293 SIETITQSSYYIGLVIMILISSFNLGAKPIATLPNTNNPAAKNGSLIVILLSFVGVPATI 352
 + S +I + + + + + PL +FT+NP + ++LL+ + P
 Sbjct: 309 AAYVRCFRSLWIAMT+VSVMVVFAPSTPLGIFTDNPDIISLIGTLLILLITILEPGR 368

-2336-

Query: 353 GTLVYTTAAWQGLGNAKLPPYTTTIGMMLIRVVVLGGLIVPSLGLLGVMMATIAINI FRW 412
 LV ++ + G+ K P Y + MW I V + YLGG+ LGL+GVW+A IAD FR
 Sbjct: 369 CNLVVSSLRAGDVKFPVYLAI VSMNGIAVPIATLILGLPLGLGLGVWIAPIADENWFRG 428

Query: 413 LFL 415
 L +
 Sbjct: 429 LLM 431

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6415> which encodes the amino acid sequence <SEQ ID 6416>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

15	INTEGRAL	Likelihood = -5.26	Transmembrane	89 - 105 (85 - 108)
	INTEGRAL	Likelihood = -4.35	Transmembrane	305 - 321 (302 - 322)
	INTEGRAL	Likelihood = -3.82	Transmembrane	161 - 177 (161 - 180)
	INTEGRAL	Likelihood = -3.82	Transmembrane	192 - 208 (189 - 208)
	INTEGRAL	Likelihood = -3.77	Transmembrane	129 - 145 (128 - 151)
20	INTEGRAL	Likelihood = -3.24	Transmembrane	242 - 258 (240 - 258)
	INTEGRAL	Likelihood = -2.81	Transmembrane	378 - 394 (377 - 394)
	INTEGRAL	Likelihood = -2.66	Transmembrane	339 - 355 (338 - 358)
	INTEGRAL	Likelihood = -2.60	Transmembrane	58 - 74 (58 - 75)
	INTEGRAL	Likelihood = -2.50	Transmembrane	32 - 48 (32 - 49)

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.3102(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 30 The protein has homology with the following sequences in the databases:

>GP:BAR06655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 119/435 (27%), Positives = 214/435 (48%), Gaps = 14/435 (3%)

Query: 9 IFSLALPSEHTENILQMLAKMVDNVLVAQIGLVAVGVSIANNISITQSLFIALGAAVSS 68
 +F+L P IE +L MANG D +++Q AV+ V ++N I+++ +F + S
 Sbjct: 11 LFAITNPFIFETILLWMLGNDUTLMLSQYSDDAVAVGVSNQLAVITVMPGFAVGTISI 70

Query: 69 LIARSGIENNNQKQLNVMAGVLQVTLFLLSVGLLSSVAGHHQVLEMLGAEASVFLVGGQY 128
 L+A+ +G + + L+ + LGLL +A +L+ + +S+ Y
 Sbjct: 71 LVAQLGAKERENAGKVAVVSGIANLIFGIVLGLLLAPGPPILKAMQLDSSLQEAITY 130

Query: 129 LSVIGMIVSLGLLTLGLAVRAQGVPIPMQVSLINVLRAIFALSTY----VNGPGL 184
 L IVGG V L+ +GAI+R+ + K M V++ +N+LN I + L I+ + G+
 Sbjct: 131 LQIVGGFVVQSLIMTAGAILSHSPTKDVVYVTIGNIIINVGNYLPIFGPFIPIVLGV 190

Query: 185 LGVWATVLSRLVGVFLICQF-----IPIQVAKRMELDKIIPDLSLPAGERLAI 236
 GVA +TV+SR +G+P++ +P +KR R ++ + +P+ACE+L
 Sbjct: 191 TGVALTIVSVRTIGLPIATILLYKRIRGLFPAYLLKRPFRVLENNLKIGIIPAGEQLS 250

Query: 237 MRAGEVLIIIGIVRFGTITAGNAIGETITQFNYPMEGLAMATITILVARQLGGGVETI 296
 A ++I + GT AL + L P ++ +A+ T IL+ Q+G ++
 Sbjct: 251 YNASQLVITYPIAMMGTEALTKVTYQNMMPVFLFAVALGQGTQILIGHQVGAQIQAA 310

Query: 297 RYIIRAFILSLMMVMGAILTYLPSLPLFTQNTDAQRSAMTIVLPSLGPADNTAGT 356
 + ++ ++ +M ++ + LL +PT ND +LL +++ P A
 Sbjct: 311 VVRCPSRLMIAMTVSVSMVVFAPSTPLIGITFTNDPILSLIGTITLLLTITILEGRACH 370

Query: 357 LVYITAVKQGLGKAKLPFYATTIGNMVIRIGLGVYGVVQVQVGLGVWATVLDNISSEPI 416
 LV + + G K P Y + MW I + + Y+G+ GLIGW+A + D R +
 Sbjct: 371 LVVSSLRAGDVKFPVYLAI VSMNGIAVPIATLILGLPLGLGLGVWIAPIADENWFRGL 430

Query: 417 LSKHFK--KYQETIF 429
 + ++ K+QE++F
 Sbjct: 431 MINRKRKKGKQEMSF 445

-2337-

An alignment of the GAS and GBS proteins is shown below.

	Identities = 219/418 (52%), Positives = 316/418 (75%)	
5	Query: 5	KELIQLAIPAKMIENILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITIYQAIPTALGASI 64
	Sbjct: 7	++I LA+P+MIENILQMLMG+VDNYLVAQ+G+VAVSGVS+ANNIITIYQ++ETALGA++ RKIPSLIALPSMIENILQMLMGVDNYLVAQGLVAVSGVS IANNIISIYQSLFETALGA+V 66
10	Query: 65	ASLLAKSLAGSKDDAISVCSQAIPLTLIGAVLIGIISIVFGQTFFKLLGTTNSVAQVGG 124
	Sbjct: 67	+SLA+A+S+ + + + + + + +TL+ LG+S+ + LG SV VGG SLLIARISIGENNNKQNTYMGVGLVQVTLTLLSVGLGLSVAGHHQVLEWLGAEASVTLGG 126
15	Query: 125	LYLAIVGGGVVTLQMLTLTGLSFLRVQCPRLPMYVSIFVNFVNAVLSGPAIFEMRYGLVG 184
	Sbjct: 127	YLAIVGG +V+LG+LT+LG+ +R QG P++PM VS+ +N LNA+ S +I+ W +GL+G QYI+SVGGMIVSLGLTSLGATVMAQGYPKIPMQVSLINVLNAIFSAISIVVWGFGLLG 186
20	Query: 185	VAVSTLIARLIGICILAKYLPKIKIHKMTWKISQAQINNALPSAGERLMAIRAGDVVIVA 244
	Sbjct: 187	VA +T++RL+G+ +L +++PIK++ KR+ + I++L+LP+NGERLMAIRAGDV+I+ VANATVLSRLVGVFLLCQFPIPKQVAKRLMRPLDKIIPDLSLPANGERLMAIRAGDVLIIG 246
25	Query: 245	IVVQLGTNVVAGNAIGETLTQFNFMPLGLIATATIIITAKYVQGNRESIETIQSSYYI 304
	Sbjct: 247	IVV+ GF +AGNAIGETLTQFNFMPLGL +ATATIIIL A+ +G I I+ ++ + IVVRFGTALAGNAIGETLTQFNFMPLGLAATATIIILVARQLGGGKVTIRIYIIEAPIL 306
30	Query: 305	GLVIMILISSFMLLAGKPLTQFTNMPNPSAIGSLVILLSFVGVGPATIGTLVYTAAMQGL 364
	Sbjct: 307	++M++++ L G L LFT N A + ++IV+L S +G PAT GLTVIA WQGL STLIMLVMGALTYLLGPSLLPLFTQNTDAQRSAIMIVLLPSLLGAPATGTLVYTAAMQGL 366
35	Query: 365	GNAKLPPYTTTIGMILIRVVVLGYLLGVFELGLLGMWMTIADNIFEMFLPKVHYHRY 422
	Sbjct: 367	G AKLPFY TTIGM+IR+ LGY++G+V++ GL+GMWMT+ DN RW L H+ +Y GNAKLPPYATTIGM+VIRIGLGYVIGVWQYGLGWMWMTVLNDTNSRWFLSKHEKKY 424
40	Identities = 46/211 (22%), Positives = 89/211 (41%), Gaps = 29/211 (13%)	
	Query: 213	MTWKISQAQINNALPSAGERLMAIRAGDVVIVAIVQLGTNVVAGNAIGETLTQFNFMPLG 272
	Sbjct: 1	M + ++I+ALPLS E ++ +V +V QHG V+G +I + + + MIYNRRKIPSLALPSMIENILQMLMGVDNYLVAQIGLVAVSGVSIANNIISIYQSLFI 60
45	Query: 273	GIATATIIITAKYVQGNRESIETIQSSYYIGLVIMILISSFML-----L 318
	Sbjct: 61	+ A L A+ +G+ N+ Q +Y G++ +L+ S L L L ALGA+VSSLIARISIGENNNK-----QLNYMAGVGLVQVTLTLLSVGLGLSVAGHHQVLEW 115
	Query: 319	AGKPLTQFTNMPNPSAIGSLVILLSFVGVGPATIGTLVYTAAMQGLGNAKLPPYTTTIGM 378
	Sbjct: 116	+ L +I G +IV L G+ ++G+V + G K+P + + + GAEASVTLVGGQYLSIVGGMIVSL----GLTSLGAIV----RAQGYPKIPMQVSLI-I 165
	Query: 379	NLIRVVLGYLLGVFELGLLGMWMTIADNI 409
	Sbjct: 166	++ + L V- GLLV AT+ + NVLNAIFSAISIVVWGFGLLGVANATVLSRL 196

A related GBS gene (SEQ ID 8071) and protein (SEQ ID 8072) were also identified. Analysis of this

-2338-

```

INTEGRAL    Likelihood = -2.18   Transmembrane   30 - 46 ( 30 - 47)
PERIPHERAL  Likelihood =  0.32           11
modified ALOM score:   2.78

```

5 *** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5564 (Affirmative) < succ>

```

10      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

```

15 ORF01629[313 - 1533 of 1878]
   EGAD[165726][TM0815 (20 - 436 of 464) conserved hypothetical protein [Thermotoga maritima]
   OWNI[TM0815 conserved hypothetical protein GP14981345[gb|AAD35897.1|AR001748.13|AR001748
   conserved hypothetical protein [Thermotoga maritima] PIR[H72331|H72331 conserved
   hypothetical protein - Thermotoga maritima (strain MSB8)
   %Match = 13.9
   %Identity = 29.4 %Similarity = 53.7
20 Matches = 120 Mismatches = 183 Conservative Sub.s = 99

```

48 78 108 138 168 198 228 258
YK*RRDTGPRCYFNLR*FVRCFFT*GGYRSTGGRSNF*NGSTYLKYARNG*RVSRFETIIKIRLP*NI*SEKETP*KFSH
25

[illegible][illegible][illegible]

45 969 995 1029 1059 1089 1119 1149 1179
TWKISAQIWNALPSAGRELMRAGDVVIVAIVVQLGTINNVAGNAIGETLQFNYPMPLGIATATIILTAKIVQGKRES
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
KWSSQKEILRVGFPTAENFVFSTGLVMANILLIAGREAYGHRGHNBSLSFMPAPGISVAITTLVRVYNGNEMGNKH

250 250 270 270 270 300 300 310 310

1205 1239 1269 1299 1329 1359 1383 1413
IETIQSSYYIGLVILMISFPMLAGKPLTLQFINNPAIKGSLVILISFVGVPATIGTIVYT--AAAGGLAKLPF
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
VLGVIRQGVLISLSPQTVGLIIIFLPEFLIRITSDQIIRKISKLPV--KIIGLFQFFADLTSTNGARLTGNILPWP

55 330 340 350 360 370 380 390

1443 1473 1503 1533 1563 1593 1623 1653
YTTTIGWLRIRVVLGIVLLGIVFGLGLGVWMTIADNIFRWFLKVVHYRIQKM*PEMWAFVFSKLIK*GLRVLFDFDI
| | : | : : : : | : | | | : | : | | |
60 IITFISINTARLPVAFVWVYFGLGLQWIGIADIIFSTLKLFLFSLGKWEKVAVLTRVRKELG
410 420 430 440 450 460

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2339-

Example 2073

A DNA sequence (GBSx2188) was identified in *S.agalactiae* <SEQ ID 6417> which encodes the amino acid sequence <SEQ ID 6418>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2200(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AD05671 GB:AB001448 THREONINE SYNTHASE [Helicobacter pylori
J99]
Identities = 161/479 (33%), Positives = 259/479 (53%), Gaps = 17/479 (3%)

Query: 14 KVTASQAILKGLADDDGLPTPTTPKVDLDFTKLQASVQYQVAKIVLSGAFFDDPTQEELD 73
K+ ++L A GGL+T F L++ SY E+ + V + + L
Sbjct: 13 KIDPIEAVINPAPKGGLYLTFET--LEWQDCLGMSYSELVHVPELLNLEIPNLLA 70

Query: 74 YCISQAYTKFELTTEIAPIVKIGDRYHL-ELPHGPTAPKDMALSILPYLLTTAAKQGV 132
+ + Y+ + APT + +R + EL+HGP+AAFKDMAL L L + A G
Sbjct: 71 SALKR-YENFDNPNPAPIPALNERLFPQELYHGPSLAFKDMALQPLASLPSNLAV--GK 127

Query: 133 ENKIVILTATSGDVTGKAAMAGFADVGTGRIIVFPKNGVSYIQELQMITQAGQNTHVVAI 192
+ K ++L +TSGDVG A + G A +P ++ YK+G S +Q+LQ+TQ N V +
Sbjct: 128 NEKYLVLVSTSGDVTGPATLEGLAGMPNIVFVCLYPKDGTSLVQKLQMTVQNASNLKVGQV 187

Query: 193 EGNFDDAQGVKEMFNLSLRKLKSOHMQLSNANSMNIGRLVPQIVYVYIYAYLAQLVSK 252
G+DDAQ ++K + + L + ++LS ANS+N GR+ QIVY+I+ + +L K
Sbjct: 188 SGDFFDAQNALNLLKDDDFNEALKARQIKLSVANSVNFGRIFAPQIVYHNGFLELYKGS 247

Query: 253 EISIGQPINFVSPTGNFNGNLAAYVASQIGLPTKILCASNDNNVLTDFPKTQTYD-KNR 311
I+ + I ++P+GNFGN L A+YA ++GL + K+ +N N+VL +T YD R
Sbjct: 248 AINSKEKITLAIPSGNFGNALGAFYAKMGNIARIKVVNTNSNDVLRFPETRIETGYDLTKR 307

Query: 312 EPKVTSSPMDILVSSNLERLPHLLGDDAETTKLMELVTGTEYALEARQANIL-ESF 370
K T SP+MDIL SSN+ER +F L G E T +L+L L YAL+ ++ +L E F
Sbjct: 308 SLWQTFSPAMDILKSNVNERALSLFG--PERTLEMAQALREKFKALKPKLALLQEHF 365

Query: 371 VAGPATEQFVELDKHLFDQYQYIEDPHTAVASAVQAYQTETKDTQPAVIVSTASPKF 430
+++ I+ ++ ++QY+ DPHTA A K ++ +TAS KF
Sbjct: 366 SCASCDEDECLKTIQEVYAEHQYLDIPHTAT-----ALNASLTKHEKTLVSATASYEKF 419

Query: 431 PCVVTKAIT-NKEEIDQPAISILNDLSGSLPKAVTDLQKAEVIRHVVPTSNMRETV 488
P A+ K+ D AA+ L+ + + DL + + H+ V+ + ++ ++
Sbjct: 420 PKTTLALNQQKKNDDDKAAETLKNYSYTPDSQRLLDLPERGKHQEVLRKINEIKSSI 478

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2074

A DNA sequence (GBSx2189) was identified in *S.agalactiae* <SEQ ID 6419> which encodes the amino acid sequence <SEQ ID 6420>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

-2340-

bacterial cytoplasm --- Certainty=0.3153 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9279> which encodes amino acid sequence <SEQ ID 9280> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAF40975 GB:AB02410 alcohol dehydrogenase, propanol-preferring
 [Neisseria meningitidis MC58]
 10 Identities = 202/282 (71%), Positives = 226/282 (80%), Gaps = 1/282 (0%)
 Query: 1 MGEIGIVVEEIGBVTSLVGDVRSIAWFFBQCHCEYCTTGRETLCRSVKNAGYSVDG 60
 +GREGIG+V+E+ +GV +L+VGDVRSIAW F+ CG CEYC TGRETL CRSV NAGY+ DG
 15 Sbjct: 60 IGHGIGILVKEVADGVKNLKVGDVRSIAWLPQSCGSCYCNITGRETLCRSVLNAGYTADG 119
 Query: 61 GMSEYAIVTADYAVKVPBGLDPAQASSITCAGVTYKAKEAGAAFGQWIAVYGAAGLGN 120
 GM+ + IV+ADYAVKVPBGLDPAQASSITCAGVTYKAIK +G PQWIA+YGAAGLGN
 15 Sbjct: 120 GMATCIVSADYAVKVPBGLDPAQASSITCAGVTYKAIKVSGVREPQWIAIYGAAGLGN 179
 Query: 121 LAVQYAKKVFNAHVAVVDINADKILQAKEVGADLVNGKEIKDVAAYIQETKGCGHVVV 180
 L VQYAKKVF AHVA+DIN DKL AKE GADL VN + +D A IQETKG H VV
 20 Sbjct: 180 LGVQYAKKVFNAHVAVDINDDKILAFKETGADLVVNAAK-EDAAKVTQETKGGAHAAVV 238
 Query: 181 TAVSKVAFNQAIDSVRAGGTVAVGLPSEYMELSIKVTLDGIRVAGSLVTRKDLLEAF 240
 TAVS AFN A++ VRAGG VVA+GLP E M+LSI + VLDGI VVAGSLVTRKDLLEAF
 25 Sbjct: 239 TAVSAAFNSAVNCVRAGGRVVAIGLPFSEMDLSIPRLVLDGIRVAGSLVTRKDLLEAF 298
 Query: 241 AFGAEGLVVPVVEKVPVDTAPQVDFEMERGLIQGRKVLDF 282
 FGAEGLVVP V+ +D AP +F EM G I GR V+D K
 30 Sbjct: 299 QFGAEGLVVPVKVQLRALDEAPAI PQEMREKGTGRVVIDMK 340

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6421> which encodes the amino acid sequence <SEQ ID 6422>. Analysis of this protein sequence reveals the following:

Possible site: 31
 35 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2356 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 263/280 (93%), Positives = 273/280 (96%)
 45 Query: 1 MGEIGIVVEEIGBVTSLVGDVRSIAWFFBQCHCEYCTTGRETLCRSVKNAGYSVDG 60
 +GHEGIGVVEEIGBVTSL+VGDVRSIAWFFBQCHCEYCTTGRETLCRSVKNAGYSVDG
 Sbjct: 76 IGHGIGIVVEEIGBVTSLKVDVRSIAWFFBQCHCEYCTTGRETLCRSVKNAGYSVDG 135
 Query: 61 GMSEYAIVTADYAVKVPBGLDPAQASSITCAGVTYKAKEAGAAFGQWIAVYGAAGLGN 120
 GMSEYA+VTADYAVKVPBGLDPAQASSITCAGVTYKAKEAGAAFGQWIAVYGAAGLGN
 50 Sbjct: 136 GMSEYAVTADYAVKVPBGLDPAQASSITCAGVTYKAKEAGAAFGQWITVYGAAGLGN 195
 Query: 121 LAVQYAKKVFNAHVAVVDINADKILQAKEVGADLVNGKEIKDVAAYIQETKGCGHVVV 180
 LAVQYAKKVFNAHVAVVDIN DKL+LAKEVGAD+ VNGKEI+DV YIQETKG HGVVV
 55 Sbjct: 196 LAVQYAKKVFNAHVAVVDINDDKLELAKEVGADLVNGKEIEDVPGVIQETKGGAHVVV 255
 Query: 181 TAVSKVAFNQAIDSVRAGGTVAVGLPSEYMELSIKVTLDGIRVAGSLVTRKDLLEAF 240
 TAVSKVAFNQAIDSVRAGGTVAVGLPSEYMELSIKVTLDGI+VAGSLVTRKDLLEAF
 60 Sbjct: 256 TAVSKVAFNQAIDSVRAGGTVAVGLPSEYMELSIKVTLDGIRVAGSLVTRKDLLEAF 315
 Query: 241 AFGAEGLVVPVVEKVPVDTAPQVDFEMERGLIQGRKVLDF 280
 AFGAEGLV VPVVEKVPVDTAP+VDFEMERGLIQGRKVLDF

-2341-

Sbjct: 316 AFGAEGLVAPVWEKVPVDTAPEVDFEMERGLIQGRKVLDF 355

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2075

A DNA sequence (GBSx2190) was identified in *S.agalactiae* <SEQ ID 6423> which encodes the amino acid sequence <SEQ ID 6424>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have a cleavable N-term signal seq.

10	INTEGRAL	Likelihood = -9.82	Transmembrane	83 - 99 (76 - 108)
	INTEGRAL	Likelihood = -7.27	Transmembrane	46 - 62 (43 - 65)
	INTEGRAL	Likelihood = -7.22	Transmembrane	187 - 203 (182 - 209)
	INTEGRAL	Likelihood = -6.00	Transmembrane	243 - 259 (229 - 262)
15	INTEGRAL	Likelihood = -4.25	Transmembrane	404 - 420 (402 - 422)
	INTEGRAL	Likelihood = -3.98	Transmembrane	120 - 136 (119 - 136)
	INTEGRAL	Likelihood = -3.88	Transmembrane	308 - 324 (307 - 324)
	INTEGRAL	Likelihood = -2.13	Transmembrane	378 - 394 (376 - 394)
	INTEGRAL	Likelihood = -1.38	Transmembrane	152 - 168 (152 - 168)
20	INTEGRAL	Likelihood = -1.17	Transmembrane	271 - 287 (271 - 287)

----- Final Results -----
 bacterial membrane --- Certainty=0.4927 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9371> which encodes amino acid sequence <SEQ ID 9372> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC17857 GB:AF026147 Yoji [Bacillus subtilis]
 Identities = 183/432 (42%), Positives = 266/432 (61%), Gaps = 1/432 (0%)

Query: 1 MKLFTPLVLIYQFANFSATTFIDSVMTGOYSQHLGAGVSTASNLWTFPFALLVGMISALVPV 60
 + + IP+ I Q TF+D+VM+G+ S LAGV+ S+LMT+ + L G+ A+ A+
 Sbjct: 15 LHLILPITITQGLSLITFLDTVMGKVSADLGAVALGSSLWTFVYTLGLAIGMAVPI 74

Query: 61 VQHLGRGNKQIETPFHQFLYLGLILSLILFLINQPIAQVIGSGLDEVLAVGRYL 120
 V Q LG R++I Q +Y+ +LS+ + +I +LG L L+ V + + +L
 Sbjct: 75 VAQLGAEKKQKIPPTVLQAVYVAALLSIAVLVIGYAAVDLILGRNLIDHVVQIAKHF 134

Query: 121 NYMLIGIMPLVLFSCRSFFDALGLTRLNMYLLMLLPNSFFNNILYKQKGMPLGCA 180
 ++ +GI FL +++ RSF D+LG TR+M+ L LP N NY+ I+GKPM LG
 Sbjct: 135 GFLLSGLIFPLFVTVYLRFSFIDSLGKTRVMTMLTSSLPNFVLYVYVIFGKGMPLGCV 194

Query: 181 GAGLGTSLTYWAIPIVILVMSLHPQIKTYHIN-TLERIKAPLIIEDIRGLGPIGLQFA 239
 GAGL ++LITW I + + + + Y I+ T+ + ++GLPIG +F
 Sbjct: 195 GAGLASALTYWCICISPFIIHKNAFFSRYGIFLWYKPSWKACKNLLKGLPIGPAVVF 254

Query: 240 EVAIFAVVGLFNAKPFSSIIIAHQAMNPFSSIMYAPFLSISTALAITISFEVGAERFQCA 299
 E +IFA V L N+ F ++ IA+HQAAMNF+SL+Y PLS+ AL I + FE GA RF+HA
 Sbjct: 255 ETSIPAAVTLNLSHFHTVTIASHQAAMNFASLLYMLPLSVSMALTIIVGFRAGAARFDA 314

Query: 300 NTYSIRIGRLTANGITSGTLFLFLERENVAAMKNSDPHVAITAQFLTYSLFFQFADRYA 359
 +YS IG + A+G + T + LERE +A MY SDP + +T FL Y+LFFQ +DR A
 Sbjct: 315 REYSLIGIMMAIGFSLFTRACILLFREQIAGMYTSDPDVRLTLQHFLLYALFFQLSDVA 374

Query: 360 APQQLIRGYKDTIKPFMIGAGSYWLCALPLAVILEKMSQLGPPFAYWIGLITGIFVCLIF 419
 AP+Q3 LRGYED SY+ LD+ ++ + LG F YWIGLI G+ +
 Sbjct: 375 APIQALRGYKDVNYTLAAAFVSYWVIGLPGVGMVGTFTSLGAGYWIGLILGALAAAGV 434

Query: 420 INQRLQKIKKLY 431

-2342-

L RL K+K Y
 Sbjct: 435 LFFRIAKLQRRY 446

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2076

A DNA sequence (GBSx2191) was identified in *S.agalactiae* <SEQ ID 6425> which encodes the amino acid sequence <SEQ ID 6426>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 23 - 39 (23 - 39)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2077

A DNA sequence (GBSx2192) was identified in *S.agalactiae* <SEQ ID 6427> which encodes the amino acid sequence <SEQ ID 6428>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3829(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC06891 GB:AB000703 hypothetical protein [Aquifex aeolicus]
 Identities = 72/213 (33%), Positives = 115/213 (53%), Gaps = 11/213 (5%)
 Query: 36 RPKILMHVCCAPCYTITLYLSQ---WADVITYFANSNIHPKDEYIRRYVTVQKRVHDFN 92
 + KIL+H+CCAF + Y L+ L + +++ YF + NIHP +EY R T++ +
 Sbjct: 3 KSKILVHICCAFDAIYFLKLRDYPESRIIGYFYDENLHPYEYRLVLELTRICKRLG 62
 Query: 93 KNTGYSVOLFSAFYFNEFFKIVHGLEEPEGGDRCKVCYDFRLDKTAEKAVELGDFYFG 152
 N + Y+ + + V G R+EPE G RC++C+D+RL+K+AE A ELG D
 Sbjct: 63 IN-----LIEGEYDLENWLERVKGYDEPERGKRCQICFYRLKSAEVAKRLGCDALT 116
 Query: 153 SALTISPHKNSQTINTIGIDVOKIYDYOYLPDLKNNKYQORSVEMCKDYDIYRCQYCG 212
 + L +SP K+ + G + K ++L D +K G Q ++ K+ +IY+Q YOGC
 Sbjct: 117 TLLMSPKKSIPLKRGAGEEATKRTGLEFLAPDYRKGGGTQEMFKLSKREETVQYQYCG 176
 Query: 213 IFGAKDQGINLIQIKKDAKAPVSDKDKGKEFPN 245
 I+G Q +I D F+ + G+E N
 Sbjct: 177 IYGLFKQKNG--KIFWDLVGLGRPGSGKEERN 207

-2343-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6429> which encodes the amino acid sequence <SEQ ID 6430>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3498 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 RGD motif: 254-255

The protein has homology with the following sequences in the databases:

>GP:AA06891 GB:AB000703 hypothetical protein [Aquifex aeolicus]
 Identities = 65/182 (35%), Positives = 106/182 (57%), Gaps = 9/182 (4%)
 Query: 39 RPSILMHVCCAPCSTTTLEYLTQF---ADITVYFANSNIHPKDEYHRRAYVTQQFVSEFN 95
 + IL+H+CCAP + Y L+ L + ++I YF + NIHF +EY R T++ E
 Sbjct: 3 KSKILVHICCAPDALYFLKLRDYPSSILIGYFDINHPHYEYRLRYLSETERIKSLG 62
 Query: 96 AKTGNVTQFLADYVFNHYVVRQVGLSEEPGGDRCKVCYDYLKDTAKAVELGQDYFA 155
 + E +Y ++ +V+G E+SPE G RC++CFDYRL+K+A+ A ELG D
 Sbjct: 63 -----INLIGETDLENLERVVKGTEDEPENKRCQICFDYRLKSAEYAKELGCDALT 116
 Query: 156 SALTISPHKNSQTINVGIDVQKYYTTKYLPSDFKNNHYERSVEMCEBYDIDYRCYCGC 215
 + L +SF K+ + G + K ++L D+K G + ++ +E +IY+Q YCGC
 Sbjct: 117 TTLMSFKKSIPLQKAGAEATKRTGIEFLAFDYRKGGQGTQEMFKLSKEREYQYQYCGC 176
 Query: 216 VY 217
 +Y
 Sbjct: 177 IY 178

An alignment of the GAS and GBS proteins is shown below.

Identities = 184/255 (72%), Positives = 219/255 (85%)
 Query: 1 MIDVENILEKMKPNQKINYDRVMQMVQMQASDIRPKILMHVCCAPCSTTTLEYLSQWA 60
 MID++ IL M PNQKINYD VMQK K W+ +RP ILMHVCCAPCSTTTLEYL+Q+A
 Sbjct: 4 MIDLQELANMNPQKINYDRVMQMAKVMKESVRPSILMHVCCAPCSTTTLEYLTQFA 63
 Query: 61 DVITIYFANSNIHPKDEYHRRAYVTQQFVSEFNKNTGVSVQFLSAFYEPNEFFKIVHGLHE 120
 D+T+YFANSNIHPKDEY+RR YVTO+V+FN TG +VQFL A Y PNE+ + V GLEE
 Sbjct: 64 DITVYFANSNIHPKDEYHRRAYVTQQFVSEFNKNTGVVLEADYVFNHYVVRQVGLSE 123
 Query: 121 EPEGGDRCKVCYDYLKDTAKAVELGFDYFSGSALTISPHKNSQTINFGIDVQKIDYDQ 180
 EPEGGDRCKVC+D+RLDKTA+KAVELGFDYF SALTISPHKNSQTIN +GLDVQK+Y T+
 Sbjct: 124 EPEGGDRCKVCYDYLKDTAKAVELGFDYFASALTISPHKNSQTINVGIDVQKYYTTK 183
 Query: 181 YLPSDLKNNKGYQSRSEVMCRDYDIYRCYCGCIFGAKDQGINLLQIKEDAKAFVSDKDGK 240
 YLESD KNN GY+RSVEMC++YDIYRCYCGC++ AK QSI+L+Q+KKDAKAF++DKD
 Sbjct: 184 YLPSDFKNNKGYRSRSEVMCRDYDIYRCYCGCVYAAKQGLDLVQVKKDAKAFMAKDL 243
 Query: 241 EEFNFIRPTFMGKSM 255
 +P +IRF++ G M
 Sbjct: 244 NDFTHIRFSYKGD 258

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2078

A DNA sequence (GBSx2193) was identified in *S.agalactiae* <SEQ ID 6431> which encodes the amino acid sequence <SEQ ID 6432>. Analysis of this protein sequence reveals the following:

-2344-

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4216(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB14809 GB:Z99118 excinuclease ABC (subunit C) [Bacillus subtilis]
 Identities = 189/333 (56%), Positives = 244/333 (72%)

Query: 1 MNSLKHGKLELLPDSGCGYLMKDKNGTIIYVGKAKNLRVKSYPHSHNTKTELLVSEI 60
 MN+ +K KL LLPD PGCVL KD+ T+IYVGKAK LNRV+SYF GSH+ KT+ LV+EI
 15 Sbjct: 1 MNKQAKKELLALPDGPGCYLMKDRQQTIIYVGKAKVLRVRSYPTGSHDAKTQRIATVEI 60

Query: 61 EDFEYIVTTSNTLEALLLEINLIQENMPKYNIRLKDDKSYPIKITNERYPRIMTQVVK 120
 EDFEYIVT+SN EAL+LE+NLII++ PKYN+ LKDDK+Y+IK+T+ER+PRL++TR VKK
 20 Sbjct: 61 EDFEYIVTSSNLEALLLENNLIKHHDPKYNVMLKDDKTYPIFKLTHREPRILIVTRNVKK 120

Query: 121 SDGTYPGYPDGGAAETIKKLLDRLFPFKKCTINPANKVCFYTHLQCCNAHTVCOITKAYW 180
 G YPGYP+ AA E K+LLDRL+P +KC+ ++VC YHLLQC A V ++
 25 Sbjct: 121 DKGRYGPYPYNQAAETIKKLLDRLYPLRKCSKLPORVCLYHLLQCLAPCVKDISSETN 180

Query: 181 DSLREDVQKFLANGDKNKI/VNGLTENKMSAAMTEFFRAEYRDILRAISLRRTKQVRIHQ 240
 L E + +FL G N++ L ERM AA +EFERA E RD I I KQ++
 30 Sbjct: 181 RELVESITRFLRGGYNEVKSLERHNEAENLEFERAKELRDQIAHISTMEKQVITVN 240

Query: 241 DMGRDVRGYPVDKGMCMQVFFVFRNGKLIQRDVNMFFTYNEPDEDFLTYYIQGYQDTRK 300
 D+ DRDVF Y DKGMMCVQVFF+R GKLI+RDV+MFP Y E +B+FLT+IQGY H
 35 Sbjct: 241 DLVDRDVFAYAYDKGMCMQVFFIIRGKLIQRDVSMFPLQEADEFELTYIQGYKNNH 300

Query: 301 FLPKBVFIPQDIDAKSVETIVGCKIVKQGRK 333
 FLPKB+ +P ID +E ++ + +P+G +
 35 Sbjct: 301 FLPKBILVPSIDGSMISQLLEINVHQPKOPK 333

There is also homology to SEQ ID 2568.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 2079

A DNA sequence (GBSx2194) was identified in *S. agalactiae* <SEQ ID 6433> which encodes the amino acid sequence <SEQ ID 6434>. This protein is predicted to be maltose operon transcriptional repressor (rbsR). Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3761(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9393> which encodes amino acid sequence <SEQ ID 9394> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAD02112 GB:AP039082 putative maltose operon transcriptional
 repressor [Lactococcus lactis]
 Identities = 64/166 (38%), Positives = 105/166 (62%), Gaps = 13/166 (7%)

-2345-

Query: 1 NGKSAIDYLYKKGKSIQFVTIDLNSEVSEERYLGYFGARKLGLNQKPKALLPDRGNPQV 60
 +G+ A+ L+ H++I FVD +EV EERY G+ A +LGL+ LLF N +
 5 Sbjct: 169 LGREAVRLLAQLNHNQISFVDTIKSTEVFEERYQGFNDNAERLGLSHD - LLFVDSNFSL 226

Query: 61 LEEFINRVKSEETALIVIGDTSVVRVQFLSPYKLKVPDDISIMTPNNSLFSHLHPYL 120
 E TAL+V+ D +S++V++ L L VP+D+S++T+KNS+F +IHPYL
 10 Sbjct: 227 RNE-----TALVVMVDVSLKVKVERLRSGQLNVFDVSLITYNNSIFGAMIHPYL 276

Query: 121 STFDINVNNGTSSVRRLIDIIKSPDKVPSSTIIVPFLERESVR 166
 +TFDI++ LG ++++++D+ + + + TII PF L RES +
 10 Sbjct: 277 TTFDIHIBQLGASAIKKILDLRDNKENLPEKTII--PFELIVRESTK 321

There is also homology to SEQ ID 5082.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2080

- A DNA sequence (GBSx2195) was identified in *S.agalactiae* <SEQ ID 6435> which encodes the amino acid sequence <SEQ ID 6436>. This protein is predicted to be 4-alpha-glucanotransferase (malQ). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2003 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAA26923 GB:J01796 amyloamylase [Streptococcus pneumoniae]
 Identities = 250/500 (50%), Positives = 329/500 (65%), Gaps = 4/500 (0%)

Query: 1 MKIRASGVLMHITSLPGDLIGTFGRBAYAFVDLIVETDQKFWQILPLTTTSFQDSPIQS 60
 MKIR SGVLMHI+SLPG GIG+FG+ AY FVDLIV T Q++WQILPL TS+QDSPIQS
 35 Sbjct: 1 MKIRQSGVLMHISLPGAYIGISFGQSAYDFVDLIVRTQRYWQILPLGATSGDSPIQS 60

Query: 61 FSAVAGNTHLIDPLLTLEGGFISKDDYQNIISPOQDPEVVYDAGLFKRRPYLEKAVKNFL 120
 FSA AGNTH ID D+L +G + D + + FG D VDI+ + RRP+LERANK F
 40 Sbjct: 61 FSAFAGNTHLIDLDILVQSLLEASDLGGVDFGSDRSEVDYAKIYIARRPLLEKAVKRF 120

Query: 121 QERATRMLSDPLQE-EKQVTDFAEPNAIKHPGNALQSWDDKAIIRREBHALAGYRQK 179
 E + F Q+ + W+ FAE+NAIKE+ F N A EW D R+ AL YR++
 40 Sbjct: 121 -EVGDVDFKFKAQDQSWLELFAEYNAIKKEYFDNLAWTEWPDADARARASALSEYR 179

Query: 180 LSEVIKYHEVTVQFFYKQWELKEVYANDKGIGLIQEMPIYVSADSVVWVWDELKFLERD 239
 L++ + YH VIOYFF+GW +LK YAND I-I+GDMPIYV+ DS +W P LFK D +
 45 Sbjct: 180 LADKLHYHVTVQYFFYQWELKAYANDMHIEVGEIMPITYVAEDSSDGNWNLHFKTDVN 239

Query: 240 KOPLAIAAGVADDFSDQGLWGNPIYNDWYHKSSDFWVIYRIOSGVWMDYLAIDHFKG 299
 + IAG P D+FS GQWGNPIYW+ + + WHI R++ K+YD +RIDHFKG
 50 Sbjct: 240 GKATCLINGCPPDEFSTQGLWGNPIYDWEAMDKQGYWIERLRESFTYIDIRIDHFRG 299

Query: 300 FSDTWIRIGVITQANDGSWQAPGPELFATIKKGLDLP+IARNIGYIDRBAERLLAGTG 359
 F YWEI TA G W PG +LFA +KE+LG+L IIAE+LG++ + L TG
 55 Sbjct: 300 FESYWEIPAGSDTAAPGWNWKGPCYKLFVAIVKELGELNIIAELDGFHDEVIKERTG 359

Query: 360 FPGKIMEFGFYDTTCSNDIPHNHYENTIAVAGTHDNEVNWGFEN-LTVBQKAYENY 418
 FPGKII++F + + SID PH N++ Y QIDHN + GW+ N + + Y Y
 60 Sbjct: 360 FPGKILQFAP-NPEDESIDSPLAPANSWYVYTGIDHNTVILGWYRNEIDDATREYMARY 418

-2346-

Query: 419 MRRLNPEPITETVLKLYATVSQTITTCMQDLLDKPADSRMMNPMTVGGNNQWRMKREDL 478
 R E + +LRT+++VS I MQDLL+ .RMN P+T+GNN WRN ++ L
 Sbjct: 419 TNRKEYETVVMHMLKTVSSVSPMAIATMCDDLELDRAARNFPSTLGGNWSWRMTBDQL 478

5 Query: 479 TENRKAPLKEITTIYNRGNK 498
 T + L +TTIY R N+
 Sbjct: 479 TPAVERGLDLITTIYRRNK 498

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6437> which encodes the amino acid
 10 sequence <SEQ ID 6438>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.85 Transmembrane 435 - 451 (435 - 451)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1341 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 313/495 (63%), Positives = 387/495 (77%)

Query: 1 MKKRASGVLMIHITSLPGDLGIGTPGRKAYAFVDPLVETDQKFWQILPLTTTSPGDSPTQS 60
 M KRAS+LMHI+SLPG GIGTPG+ A+ FVDPL ET Q +WQILPLTTTSPGDSPTQS
 25 Sbjct: 1 MNKFASGILMIHISLPGKFGIGTPGRKAFEPVDPLASTIKQTYWQILPLTTTSPGDSPTQS 60

Query: 61 FSAVAGNTHLIDFOLLTLEGFISKDDYQCNISPGQDPVVYDAGLEKRRPVLEKAVKNFL 120
 FSA+AGNTH ID+LL + + D +I+PG +PE VDYA LP+ RRP+LEKAV+ P+
 30 Sbjct: 61 FSAIAGNTHFIDPELLVDELLEADLDCITTPGTPNPAVDYAGLPQVRPLLEKAVRAFY 120

Query: 121 QBERATRMLSDFLQBEKNVTDFAEFMAIKHFPGNKALQEWDDKAIIRSEELAGYRQKL 180
 E+ L F W+TDFAEFMA+KE+F NKALQ+WDD+ +I+R+E+L YR+ YL
 Sbjct: 121 AEQENVCLEAFETASSWLTDFAEFMAIKHFPGNKALQEWDDETVYIKRQEDSLNNYRELL 180

Query: 181 SEVIKYHEVTQYFFYKQMFELKEYANDKGIGIQIDGMPYVSADSVIEWTMPELFKLRDK 240
 ++ I YH+V QYFFY+QW LK YAN KGI+IIGDMPYVSADSVIEWTMPELFK+D DK
 35 Sbjct: 181 AKKITVHKVCQYFFYQMSALKTYANHKGITIGDMPYVSADSVIEWTMPELFKVDSDK 240

Query: 241 QPLAIGAVPADSPSDGQLGMPIYNMIDYHKESDPFWMIYRIQSGVMYDYLRIHDFKGF 300
 +PL IAGVPAD PS+DQQLGMNP YNW E++S+F WHIYRIQ K+YD LRIDHFKGF
 40 Sbjct: 241 KPLFIAGVPADGPSDEDGQLGMPIYNMAREKSNFANWHIYRIQSFYLDQLRIDHFKGF 300

Query: 301 SDYWEIRGDYQTDNGSQWQAPGPELFAITKEKLGDLPIIAENLGYIDRAERLLAGTGF 360
 SD+WEI +TA +G W AFG LE+ +E LG+LPIIAENLGYIDE+AE+LLA TGF
 45 Sbjct: 301 SDFWEIPAGDKYARNGHWSARGDALFSAVREALGELFIIAENLGYIDRAEKQLLASTGF 360

Query: 361 PQMKINEFGFYDIGNSIDIPHYNTENTIYATGTHDNEVINGWFENITVEKAYAEYMR 420
 PQMKI+EPG +D T SID+PH Y N +AY GTHDNEV+NGW++NL+ BQ + NY+
 50 Sbjct: 361 PQMKILEFGLFDITSSQSIDLPHYDKNCAVYTGTHDNEVINGWYDNLSESGVHFVNHTLH 420

Query: 421 RLPNPEPITETVLKLYATVSQTITTCMQDLLDKPADSRMMNPMTVGGNNQWRMKREDL 480
 + +E IT+ +LRT++A+V T I C+QDLLDK SRMMNPMT+GNNQWRM +L +
 Sbjct: 421 KEADESITKAMLETTIPASVCDTALLCIQDLLDKDGKSRMMNPMTIGNNQWRMLDGLN 480

Query: 481 NRKAPLKEITTIYNR 495
 + K +L +T +Y R
 Sbjct: 481 DEKDYLYIYDLYGR 495

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 60 vaccines or diagnostics.

Example 2081

A DNA sequence (GBSx2196) was identified in *S.agalactiae* <SEQ ID 6439> which encodes the amino acid sequence <SEQ ID 6440>. This protein is predicted to be glycogen phosphorylase (maIP). Analysis of this protein sequence reveals the following:

5 Possible site: 40
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2678 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC00218 GB:AF008220 glycogen phosphorylase [Bacillus subtilis]
 Identities = 297/776 (38%), Positives = 452/776 (57%), Gaps = 41/776 (5%)

20 Query: 13 GKVLSELNREIYVELLNFKERAA-----KSNSSQKRVYIISAEFLIKLSNNL 65
 GK + + Y L N V+E +A KS+++S ++ YY+S EFL+G+LL NL
 Sbjct: 21 GKSFKDAKLDQYKTLGNMVRVEISADWIEKSRSGKQTYLSIEFLGQLEQNL 80

25 Query: 66 INLGIYKDVKKELELVGKSIASIEDVPEPSLNGGLGRASCDFDSISLGINGBGVGL 125
 +NLG+ V+ L+ +G ++ EI +E + LGGISGRLA+CF+DS++SL + G G+G+
 Sbjct: 81 MNLGVRDVVRAGLKEIGINLEIILQIENDAGLNGGLGRLAACFLDSLASNLGKMGVI 140

30 Query: 126 NYRGLGFQKQVFRNNQAEANYWIEN-NSWLVPD-DISYDVFF-----RDFILKSRLL 175
 Y GLF+Q + Q W+N N W V D + DVFF + L R
 Sbjct: 141 RYKHGLFQKIVDGHQVLEPQMLQKGNVWVRNADQAVDPFVGEVHTKSGRLHFRH 200

35 Query: 176 DR-----IDVLGYKKDTKNYLNLFIDIGLNYLRIKGITFDKTEIKNNLTLFLYP 225
 ++ I ++G+ T N L L++ + Y G + ++ FLYP
 Sbjct: 201 EQATIVAVPYDIPIGYETGVTNLRNAB--PYAHYHGGNLSYKRETAVERMDFH 258

40 Query: 226 DDSDKNGELLRIYQCYFMVSNAAQLLIDEAIEBGSNLDLAETAYVQINDTHPSKVIPEL 285
 DD+ G++LR+ QQYP+V + + +++ + +L L + + INDTHP++ +PEL
 Sbjct: 259 DDTDEGKILRLKQQYFLVCAKSLKSVNNYKTRKTHSLGLHKVSIHINDTHPALAVPM 318

45 Query: 286 IRLLTEKHGFPEDEAVSVVRNMVGYNHTTLAEALEKWPLEYLNEVVHVLTVTIKKLDQM 345
 +R+L ++ ++EA + + + YTNH L+EALEKWP+ ++P + II++++
 Sbjct: 319 MRILLDEENMSWESARHITVHTIISYTNHTTLSEALEKWPFIHLFKPILLPMYIMTEINER 378

50 Query: 346 IRE-----EQTNPEVQIIDEAGRVIMAHMDIHFSTSVNGVAALHTSIKNSKLVFY 397
 + E I G V MAH+ I S SVNGVA +H++ILK E++ F+
 Sbjct: 379 FCRAVWEKYPGDWKRITENMALTAGVVKMAHLATVGSYSVNGVAKIHSILKEREMDFH 438

55 Query: 398 DIYPDKFNNTKNGITFRNLEFANQQLADYLKELIGDSYLTDTQLEKLLTYADSNEVHD 457
 ++P++FNNTKNGI RRWL AN L+ + E IGD ++ L +L YA +
 Sbjct: 439 LLFPNRFNNTKNGIAHRRNLLQANQGLSALITTEAGDEWVKQESLIRLEPYATDPAPIS 498

60 Query: 458 KLAAIKFKNKLAKRYLKENKGIELDEYSIIDIQIKRFPEYKQKQNNALYVIRKYLEIKR 517
 + K K K L + G+ ++ SI D Q+K R E YKRQ +N L+++ Y +K
 Sbjct: 499 QFQNNKSKKKQELADLIFCTAGVNVNPEISIDVQVKRLFAYKQQLNVLHIMYLRNLKE 558

 Query: 518 GH-FPSKKLTIVPGKAAAPYTTAQDI IHLILCLSELINNDPEVKNYLVNHLVENVYTV 576
 F T IFG KA+P+Y A+ II LI ++E +N DP V++ + V +BNY V++
 Sbjct: 559 DSGFSIYPTFTIPGAKASPSYYAKKIILKTHSVAKVNYDPAVKQLIKVPLRNRYVSM 618

 Query: 577 AEKLIPTADISBQISLASKEASGTGKMKPMLNGALTLTGMDGANVEIASLAGKRNITYFG 636
 AE++ PA+D+SBQIS ASKEASGTGKMKPMLNGALT+GT DGAN+EI E G +ITYFG
 Sbjct: 619 AERIFPASDVSBQISTASKKEASGTGKMKPMLNGALTIGTHDGANIEIRVGPDCITYG 678

 Query: 637 KDSDTIINLYETSGYRSKDYDKKVIKREAVDFIISDDIVLSNAERLKRHLDELV-GKD 695
 +D +++ E GYRS++Y D+ IR+ D +I+ G A+ + + D L+ D
 Sbjct: 679 LKADRVLSYQRNGGYRSREYQKDRIRIQVADQLINGFE--GRADEPESIDSLIFHND 736

-2348-

Query: 696 WFMTLIDLKEYIAVBQVLADYDYBSWNKKVHNHTAKAGFPSSDRTEBQYNQDIW 751
 + L D Y +E+ ADY + W++ I NIA +G+PSSDRTI +Y +DIW
 Sbjct: 737 EYFVLKDFSSYADQERIQADYRERRKWEHSIVNIAHSQYPSDRTEYAKDIW 792

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6441> which encodes the amino acid sequence <SEQ ID 6442>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.71 Transmembrane 538 - 554 (538 - 554)

----- Final Results -----

bacterial membrane --- Certainty=0.2084 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 629/754 (83%), Positives = 696/754 (91%), Gaps = 2/754 (0%)

Query: 1 MTRNFTTYVGQQ-GKVLSELTNEEIVVELLNFVKERAAKSNKSSQRKVYISAEFLIGK 59
 MTR FT YV + GK L++ NNEEIV+ LILNFVKER+ K+KNS++RKVYISAEFLIGK
 Sbjct: 1 MTR-FTYVETLKGSLTQASNNEEIVLSLILNFVKERASHKAKNSAKRKVYISAEFLIGK 59

Query: 60 LLSNNLNLNLGIYKIVKKELELVGKSIAREIDVPEPSLNGGLGLRLASCFDISISLGIN 119
 LLSNNLNLNLGIYK+K+EL GKSIAR+EDVE EPLSNGGLGLRLASCFDISI+SLGIN
 Sbjct: 60 LLSNNLNLNLGIYKIKETIAAAGKSIAREIDVELEPSLNGGLGLRLASCFDISITAGLGIN 119

Query: 120 GEGVGLNYHCGLFQKVFNNQAEANYWIENNSMLVPTDISYDVFPFDRLKSLRLDRID 179
 GEGVGLNYHCGLFQKVF++N+QAE N+WIE++SMLVPTDISYDVFP++PTLKSRLDRID
 Sbjct: 120 GEGVGLNYHCGLFQKVFKNQAEAPFNWIEDDSMLVPTDISYDVFPFNPTLKSRLDRID 179

Query: 180 VLGYKDTKYNILNLFIDGLDYNLEBKGITFDKTEIKKMLTFLYLPDSSDKIGELLRIYQ 239
 VLGY+DTKYNILNLFDI+G+DY LI+ GI+FDKT+I KMLTFLYLPDSSDKIGELLRIYQ
 Sbjct: 180 VLGYKDTKYNILNLFIDBGVGLKDGISFDKTIKKNLTLFLYLPDSSDKIGELLRIYQ 239

Query: 240 QYPMVSNAAQLLIDEAIERGNSNLHDLAYATVQINDTHPSMVIPELIRLLTEKHGFDFE 299
 QYPMVSNAAQL+IDEAIERGNSNLHDLA+YATVQINDTHPSMVIPELIRLLTEKHGF+DFE
 Sbjct: 240 QYPMVSNAAQLLIDEAIERGNSNLHDLAYATVQINDTHPSMVIPELIRLLTEKHGFDFE 299

Query: 300 AVSVVRNMVGYTNHTLAALEAKMPLYELNEVVPHLVITIIKKLDQMIREBQTNPEVQIID 359
 AV+V+NMVGYTNHTLAALEAKMP YLNEVVPHLVITII+KLD ++R E ++P VQIID
 Sbjct: 300 AVAVVNMVGYTNHTLAALEAKMPTAYLNEVVPHLVITIIKKLDALVRSVEVDVAVQIID 359

Query: 360 EAGRVMHMDIHFSTSVNGVAALTEILKNSLEKVFYDIYDFKNFKNTGKITFRWLEF 419
 E+GRVHMDIHF+TSVNGVAALTEILKNSLEK FDI+Y+KFNFKNTGKITFRWLEF
 Sbjct: 360 EGRVHMDIHFSTSVNGVAALTEILKNSLEKAFYDLYPEKFNFKNTGKITFRWLEF 419

Query: 420 ANQLADLYKELIGDSYLTATLCKELLYADSNVHDKLAAIKFKNKALKRKYLNKNG 479
 ANQLADY+KELIGD YLTAT+LEKL+ +AD VH KLA IKF NKLAKRKYLNK
 Sbjct: 420 ANQLADLYKELIGDEYLTATLCKELMA+ADKAVHAKLAJIKFKNKALKRKYLNKDK 479

Query: 480 IELDEYSIIDTIKRPHEYKRCQMNALYVIHKYLEIKRGHPSPKRLTVIYGKKAAPAYT 539
 IELDE+SIIDTIKRPHEYKRCQMNALYVIHKYLEIK+G+ P RK+TVIYGKKAAPAYT
 Sbjct: 480 IELDEHSIIDTIKRPHEYKRCQMNALYVIHKYLEIKKGNLPRKRTIVIYGKKAAPAYT 539

Query: 540 AQDIHILCLSELINNDPEVNKYNVHLVENNVTVAREKLIPATDISEQISLASKEAG 599
 AQDIHILCLSELINNDPEV+ YNVHLVENNVTVAR LIPATDISEQISLASKEAG
 Sbjct: 540 AQDIHILCLSELINNDPEVSYLVNHLVENNVTVAREHLIPATDISEQISLASKEAG 599

Query: 600 TGNMKFNLGALTGTMDGRNVIAELAGKNLYTFGKDSDTIINLYETSGTRKDYIDK 659
 TGNMKFNLGALTGTMDGRNVIAELAG ENLYTFGKDSDTIINLY T+Y +KYID
 Sbjct: 600 TGNMKFNLGALTGTMDGRNVIAELAGKNLYTFGKDSDTIINLYATATVAKDYIDN 659

Query: 660 DKVIREAVDFIISDDIVSLGNARLKRHLDELGVKQWFMTLIDLKEYIAVBQVLADYD 719
 I+ AV+FIIS ++++ GN ERL L+ EL+ KQWFMTLIDE+HYI VKR++LADYD
 Sbjct: 660 HPAIKAAVNFIIPELLAFGRERLRLRYKELISKQWFMTLIDLEYIEVYKRLADYD 719

-2349-

Query: 720 YESWNKKVHNIKAGFFSSDRITIEQYNQDIWHS 753
 + W KV+HNIKAGFFSSDRITIEQYN+DIWHS
 Sbjct: 720 QDLWHTKVHNIKAGFFSSDRITIEQYNEDIWHS 753

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2082

A DNA sequence (GBSx2197) was identified in *S. agalactiae* <SEQ ID 6443> which encodes the amino acid sequence <SEQ ID 6444>. This protein is predicted to be glycerol-3-phosphatase transporter (gipT). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.88	Transmembrane	339 - 355 (333 - 359)
INTEGRAL	Likelihood = -8.01	Transmembrane	432 - 448 (426 - 450)
INTEGRAL	Likelihood = -6.74	Transmembrane	92 - 108 (91 - 127)
INTEGRAL	Likelihood = -6.69	Transmembrane	194 - 210 (190 - 214)
INTEGRAL	Likelihood = -3.77	Transmembrane	367 - 383 (364 - 385)
INTEGRAL	Likelihood = -2.81	Transmembrane	111 - 127 (109 - 127)
INTEGRAL	Likelihood = -2.28	Transmembrane	407 - 423 (406 - 424)
INTEGRAL	Likelihood = -2.02	Transmembrane	165 - 181 (165 - 182)
INTEGRAL	Likelihood = -0.64	Transmembrane	29 - 45 (29 - 45)

----- Final Results -----

bacterial membrane	---	Certainty=0.5352(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA044575 GB:U28354 I8629 ORFB fused with sequences similar to E. coli gipT and UhpT proteins, Swiss-Prot Accession Number P08194 and P09836; Method: conceptual translation supplied by author [Shig]

Identities = 174/321 (54%), Positives = 241/321 (74%), Gaps = 4/321 (1%)

Query: 109 GVPSVITSIWLFPTIMYLINGMLQGMGYPPGARTLVVWYDNKERIKYATIMNLSHNVFGGA 168
 GV P V + + + YL+NGM+QGMGYPPGA+TLV+VY++ERI +AT+MNLSHN GGA
 Sbjct: 12 GVGP-VCSELHAPSTYLLNGWQGMGYPPGARTLVVWYEHRRERISWATMNLSHNVGGA 70

Query: 169 LAPILTVGLALGNDSLNQRAAYWFFGVVACILAVLVYPLQEDTFESIGLPIREYHK 228
 +AP+L G + G +L+ ARAA+ FPGV+ ++VL+YF+Q D P S+GLPIRE+
 Sbjct: 71 LAPVLIGSPGPGGSDALHAKAAFPFGVLCNMMSVLIYFIQWDRFVSGLPIREYHK 130

Query: 229 EQYTNVVDSSDILEEPEVLGMGEIKKYILPNTKLMASLYSIFVYILRYGIVSWVPKFL 288
 ++ E+ L + +II+G+I+ N KI++ +Y FVYILRYGIVSW PKFL
 Sbjct: 131 NVVSHPAKGR--EQGPRISIPDIRKHIIIRNNKLIYCCYISFVYILRYGIVSWAPKFL 187

Query: 289 ATSVQDGGKGITATAGMGPSLFEIGGILGMLTAGYLSAKVFRNSKELTNVAFIVAVLL 348
 + S+ GK+ A MGS S+PRIG+ GML AGYLS ++F+NSKPLTN FL+ I+L
 Sbjct: 188 SDGLDVGCKDMGKLAMGCSVSFEIGGVAGMLLAGYLSVRLFRNSKELTNVAFIATITIL 247

Query: 349 LAAYWFIAPGFQYMAIDFILLGLGASTYGPVWVGLYAMELVKPAAGAASGLITVPSY 408
 L AYW++P+G +Y+ L+ IL+ LG ++YGPVM +GLY+MELVPEK AAGAASGL+GTPSY
 Sbjct: 248 LIAYWVPSGNEYLWNTYILLGLAVYGPVWVGLYAMELVKPEAAGAASGLSTVPSY 307

Query: 409 VGGATATATLAIIGIHDPGNG 429
 + G+ +ATL +G++D+ GNG
 Sbjct: 308 IPGSIVATIGMGLVVDYLGNG 328

-2350-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6445> which encodes the amino acid sequence <SEQ ID 6446>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence
5  INTEGRAL Likelihood = -12.37 Transmembrane 185 - 201 ( 175 - 208)
   INTEGRAL Likelihood = -9.13 Transmembrane 114 - 130 ( 90 - 134)
   INTEGRAL Likelihood = -7.75 Transmembrane 322 - 338 ( 320 - 345)
   INTEGRAL Likelihood = -6.79 Transmembrane 421 - 437 ( 419 - 439)
10  INTEGRAL Likelihood = -6.37 Transmembrane 91 - 107 ( 90 - 113)
   INTEGRAL Likelihood = -5.36 Transmembrane 163 - 179 ( 161 - 181)
   INTEGRAL Likelihood = -5.20 Transmembrane 350 - 366 ( 347 - 371)
   INTEGRAL Likelihood = -4.41 Transmembrane 23 - 39 ( 22 - 41)
   INTEGRAL Likelihood = -3.77 Transmembrane 257 - 273 ( 249 - 273)
   INTEGRAL Likelihood = -1.33 Transmembrane 61 - 77 ( 61 - 77)
15  INTEGRAL Likelihood = -1.28 Transmembrane 383 - 399 ( 383 - 399)
   INTEGRAL Likelihood = -0.90 Transmembrane 299 - 315 ( 299 - 315)

----- Final Results -----
bacterial membrane --- Certainty=0.5946(Affirmative) < succ>
20  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF96050 GB:AE004355 glycerol-3-phosphate transporter [Vibrio cholerae]
Identities = 128/438 (29%), Positives = 215/438 (48%), Gaps = 17/438 (3%)
25  Query: 1 LPMEDYNNKREP-EKFTQFLRQKVFFVFAFF-GYVCAYLVRRNFKLMSNTIMVQNGMDK 58
    LF + +R P +K R + F+ F GY YL R NF L + + +G + +
    Sbjct: 21 LFKPAATQRLPSDKVDGVSRLRWQLFICIFVGAGYVLGRNIFSL-AMFYLIBEQPFR 79

30  Query: 59 AQIAILLCLTYSYGLAKFYMGALGDRVSLRKLFSISLGASALICILIGFF---NSSM+V 115
    + + L +++YGL+KF MG + DR + R S L SAL+ GF S+
    Sbjct: 80 GDLGVALGAVSIAYGLSKFLMGNVSDRSNPRYFLSGLLLGALVWFCFPMFWATGSITA 139

35  Query: 116 LGILLVLCGVVQALAPASQMIANYFFNKTTRGAGLWNI SQNMGSALEPLTIALTSM 175
    + ILL L G QG PA + + + K RG + + VN+ N+G L I + +
    Sbjct: 140 MFILLFLNGWFGQXNPACGRIMVHMSRGERGEIVSVVNVAINVGGGL---IGPIFL 195

Query: 176 GLVVFANGNILLAFILIPGVLVFLFALCCMKLGDNFSEGLDLSRTMYODAGESAVASE 235
    GL + N + AF +P L A+ W + D P+S GL + D + S E
40  Sbjct: 196 GLWM-FNDWRTAFYVPAFFAVLVAVFTWLVNRDTPQSCGLPPIREYKNDIPDDYDKSE 254

Query: 236 EKINLSYQILWIKYVFCNPSLLWAAVNVALYFVRFOIEDMPITYLGSQVNMGRRAHIFA 295
    + + + + +KCVF N L +A N +Y +R+G+ DW P+YL + + + +A
    Sbjct: 255 NE--MTAKEIFFKCYFNNKLKLSIAIANAPVYLIRYGLDWAAPVYLKEAKHFTVDKSSGA 312

45  Query: 296 ISMLEWVAIPGSLVFAWLAIR-YENKMAKVGAIGFLVLAIVFYVERLITATGAPNYFLLL 354
    + EW IPG+L+ W++ + + + G + + + + + V VY G P +
    Sbjct: 313 YFLYEWAGIPTGLGQWISKVKFGRAPAGILEMVLVTLAVLVY-WFNPAQNPADVMAA 371

50  Query: 355 VIAGILGSLTYGPOLIVNLTINFEVINVAGTAFGVGVITAYLIGNWGNMMLIADG 414
    ++A +G LIYGP +++ + + P AGTA G G+ YL G + AN ++ D F
    Sbjct: 372 LVA--IGFLTYGPVNLIGLYALELAPKKAAGTANGLTGLFGYGGAVANAILGYTVDFH 429

Query: 415 GWFKSYIVVAALSAPSAV 432
55  GW ++V+ A S +
    Sbjct: 430 GWDGGMVLVASCVLVL 447

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 117/439 (26%), Positives = 203/439 (45%), Gaps = 27/439 (6%)
60  Query: 23 KYPRYRVQVLISIFVGYMGYFVPRNFTSILGLNMS---ATEIGILTCSAYIAGLSK 78
    ++ R + V F GY+ Y VRN ++S + + +I + +
    Sbjct: 17 QFLRRQKVFFVFAFPGYVCAYLVRRNFKLMSNTIMVQNGWKAQIALLSCLTYSYGLAK 76

```


-2351-

Query: 79 FISGLISDESNSKIFLPVGLFLTLGVNVLIGVIPSVITSIWLFITMYLINGLQGMGYPP 138
 F G + D + + + L + LIG S S + + I + + G + QG P
 Sbjct: 77 FYMALGDRVSLRLKLFSLIGASALICILIGFPHS ---SMVVLGILLVLCGVVQALAPA 133

5 Query: 139 GARTLVVYNDKERIKYATINMLSHNFGGAIAP-----LTGVGLAAGNSINQARAAYW 194
 + ++ NK R WM+S N G A+ P+ LT +GL + N ++ A+
 Sbjct: 134 SQAMIANYPFNKTRGGAIAGNNISQNMGSALLPITLALLTSMGLVVPANGNI---LLAFL 190

10 Query: 195 FPGVVACILAVLVYFLQEDTPESIGLPPIETTHKQYINVDSSDILEEPEVLGMGEIHK 254
 PGV+ L A+ + L D PES GL + + + V S EE L ++I
 Sbjct: 191 IPGVLVFLFALCCWKLGDNPESBGLSLKIMYGDAGSEAVASE---BKKHNLSTYQLIL 247

15 Query: 255 KYILPNTKLMWASLYSIPVYLLRYGIVSWTFKPLATSVDQGGKITATAGNMGPSLFRIG 314
 KY+ N L+ + ++ +Y +R+GI W P +L+ I S+ E
 Sbjct: 248 KYVFCNPSILLVAENVVALYFVRFGIEDWMPITLSQVANMSEAHIHFA----ISMLENV 302

20 Query: 315 GIIGMLTAGTSLAKVFKNSKPLTNVAFIVVAILLAAYWFIAG-POYMALDFIILLG-L 372
 I G L +L+ + + + V+A ++ G P Y L + ++ G L
 Sbjct: 303 AIPGSLVPAWLAVRYPNMAKVAIGLFLVLAALVFVYERLTATGAPNYFL--LVIAQIL 360

25 Query: 373 GASLYGPMVGVLYAMELVFKAAAGASGLIGTFSYVGGATIATLAIIGIIDHPGQGVAF 432
 G+ IYGP ++V + + VP AG A G G +Y+ G A + I+ D FGW ++
 Sbjct: 361 GSLYGPQLIVNLTINFTVFLNVAITAIGFVGVYATLYLNGANWMLPIADGFGWFNSY 420

30 Query: 433 IIF-GISGFAIVCTILSR 450
 I+ +S F+A+ +L++
 Sbjct: 421 IVVAALSAPSAVGYLLAK 439

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 2083

A DNA sequence (GBSx2198) was identified in *S.agalactiae* <SEQ ID 6447> which encodes the amino
 acid sequence <SEQ ID 6448>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3202 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6449> which encodes the amino acid
 sequence <SEQ ID 6450>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4473 (Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/100 (54%), Positives = 67/100 (67%)

55 Query: 1 MTYELCLRYGTYPFLRFVDAWDEINTAPAFITEDKKLLLELLEEVNITFLPHELTITICSFH 60
 MTYELCLRYGTYPFL VDR+ E P F I ED+ L LE +N LFH+LF+TIE FH
 Sbjct: 1 MTYELCLRYGTYPFLSRVDAYWGRDQNPPTTFIQEDRLLCHEKLETNHILPHDLFVTIESQFH 60

-2352-

Query: 61 YIGHDFPEKRAKITQIYHVITIEHLSITHYPEYDIKTESLIM 100
 Y+G + PEKRA+I +Y + L Y +Y IKIE+ L+
 Sbjct: 61 YVGFNMPKEKRAQIRILYQEVATI LKBYKDYPIKIEITFL 100

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2084

A DNA sequence (GBSx2199) was identified in *S. agalactiae* <SEQ ID 6451> which encodes the amino acid sequence <SEQ ID 6452>. Analysis of this protein sequence reveals the following:

10 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2369 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAB81912 GB:U92974 unknown [Lactococcus lactis]
 Identities = 213/322 (66%), Positives = 260/322 (80%), Gaps = 5/322 (1%)

Query: 1 MSEKIRVLLYYKYVSIENAEYAAKHLEFCKSIGLGRILIADEGINGTVSGDYETTQKY 60
 M++ RVLLYY+YV IE+ E +A KHL CK +GLKGRIL+ADEGINGTVSG E T Y
 Sbjct: 1 MTQDYRVLLYYQYVPFIEDGETFAQKHLADCKELGLKGRILVADGEINGTVSGTIBQTNAY 60

25 Query: 61 MDVHSDERFADLNFKIDEEHQAFKRMFVRYKISIVHLGLEDDNNPDSIDINPLETTGEYL 120
 M+ + +D RF+ FKIDE Q AF+RM VRY+ E+V+L LRD I+NPLE TG YL
 Sbjct: 61 MELMKNDPRFSSTIFKIDEAEQNAPKMHVRYRPELVNLSLED-----DVNPLEITGAYL 115

30 Query: 121 NPKQFKKALLDEDTVVLOTENDYEDLGHFRGAIRPDIRNRELPGWVNDKDKFMEKRV 180
 +PK+F+EA+LDE+TVV+D RNDYE+DLGHFRGAIRP+IR+FRELPQW+RDNK++FMEKRV
 Sbjct: 116 DPKRFREAMLDENTVVIDARDNYPEDLGHFRGAIRPEIRSFRELPGWIRDNKQPMKRV 175

35 Query: 181 VVYCTGGVRCEKPSGMMVREGFKDVQQLHGSIATYGKDEPVQGLNDGAMYVFDRIISVP 240
 + YCTGG+RCEKPSGW+VREGFKDVQQL GGIATYGKDEPVQGLNDG MYVD RI+VP
 Sbjct: 176 LTYCTGGIRCEKPSGMLVREGFKDVQQLGSIATYGKDEPVQGLNDGQMYVFDRIISVP 235

Query: 241 INHVNPTVISKDYFDGTPCERYVMNCANFPCKNQIFASEENBAKYVRGCSPECAHRENNRY 300
 IN ++ +D+FDG+PCERY+NC NF CN+Q+ ASEENBAKY+ CS BCR H NRY
 Sbjct: 236 INQKEHVTVGRDWDGSPCERYINCGNFCNRCMLASEENBAKYLGACSHCRVHPNNRY 295

40 Query: 301 VQENGLSRQEWAEERLEAIGESL 322
 ++ + LS QE ERL + + L
 Sbjct: 296 IKAHQLSNQEVQERLALIEKDL 317

45

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6453> which encodes the amino acid sequence <SEQ ID 6454>. Analysis of this protein sequence reveals the following:

50 Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2443 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

Identities = 321/324 (99%), Positives = 323/324 (99%)

Query: 1 MSEKIRVLLYYKYVSIENAEYAAKHLEFCKSIGLGRILIADEGINGTVSGDYETTQKY 60

-2353-

MSKIRVLVLYYKYVSIENNA+KYAAKHLEPCSKISGLKGRILIADBGINGTVSGDYETTQKY
 Sbjet: 1 MSKIRVLVLYYKYVSIENNAQBYAAKHLEPCSKISGLKGRILIADBGINGTVSGDYETTQKY 60

 Query: 61 MDVWSDRERPADLWPKIDRENNQAARPMFVRYKKKEIVHGLSDNNFSDINPLETTGEYL 120
 MDVWSDRERPADLWPKIDRENNQAARPMFVRYKKKEIVHGLSDNNFSDINPLETTGEYL
 Sbjet: 61 MDVWSDRERPADLWPKIDRENNQAARPMFVRYKKKEIVHGLSDNNFSDINPLETTGEYL 120

 Query: 121 NPKQFTEALLDEDTVVLDTRNDYSEYDLGHFRGAIRPDIRNFRELPCWVRNKKDKPMSEKV 180
 NPKQFTEALLDEDTVVLDTRNDYSEYDLGHFRGAIRPDIRNFRELPCWVRNKKDKPMSEKV
 Sbjet: 121 NPKQFTEALLDEDTVVLDTRNDYSEYDLGHFRGAIRPDIRNFRELPCWVRNKKDKPMSEKV 180

 Query: 181 VVYCTGGVRCCKPSGMMVREGFKDVGLHGGIATYTKDPEVQGBELMDGAMVFDRIISVP 240
 VVYCTGGVRCCKPSGMMVREGFKDVGLHGGIATYTKDPEVQGBELMDGAMVFDRIISVP
 Sbjet: 181 VVYCTGGVRCCKPSGMMVREGFKDVGLHGGIATYTKDPEVQGBELMDGAMVFDRIISVP 240

 Query: 241 INHVNPVTSKDYFDGTPCERYVNCANPPCMKQIFASENEAKTVRGSCSPCRAHERNRY 300
 INHVNPVTSKDYFDGTPCERYVNCANPPCMKQIFASENEAKTVRGSCSPCRAHERNRY
 Sbjet: 241 INHVNPVTSKDYFDGTPCERYVNCANPPCMKQIFASENEAKTVRGSCSPCRAHERNRY 300

 Query: 301 VQENGLSRQEWAEKLEAIGESLP 324
 VQENGLSRQEWAEKLEAIGESLP+
 Sbjet: 301 VQENGLSRQEWAEKLEAIGESLP 324

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2085

A DNA sequence (GBSx2200) was identified in *S.agalactiae* <SEQ ID 6455> which encodes the amino acid sequence <SEQ ID 6456>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have an uncleavable N-term signal seq

 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC83954 GB:L47548 putative [Bacillus subtilis]
 Identities = 54/192 (28%), Positives = 89/192 (46%), Gaps = 14/192 (7%)

 Query: 5 QTIIIGAGAGIGPGSAMRQLGUNTFLIIEKHIGESFLRWPRTTQFITPSFTTNGFGFP 64
 + IIIG G G+ +++G+ + L+IEKG+ S +P F + S
 Sbjet: 5 KALIIIGGPGCLSAATHLKOIGI-DALVIEKGNVNSIYNYPTHQTFPSSSEKLE----- 58

 Query: 65 DLNAVIPDTPSPAFSEKHLGSGVEYARYLQVAAHYNPLIQNETSLVSDIK-RDSLFIK 123
 I D AF E ++ Y + V N+ + V + K +++ FVI+
 Sbjet: 59 -----IGDV--AFITENKKVRIQALSYRRVVKRKNIRVNAPAMVRKVITYKQNTFVIR 111

 Query: 124 TSKGDSADYLIMATGEFQNPNTIDIKGADLGMHYGQVDFHIKSDNPFIIIGNNSACD 183
 TSK ++ Y I+ATG + +PN + + G DL + H D + +TGG S+ D
 Sbjet: 112 TSKETITPTTCIIATGYDHPNNMGVPGKDLKVFHYFKRGHPYFDKDVVVIGKNSVD 171

 Query: 184 ALTHLVYLGQIV 195
 A LV G + V
 Sbjet: 172 AALELVKSGARV 183

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2354-

A related GBS gene <SEQ ID 8973> and protein <SEQ ID 8974> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop Possible site: -1      Crend: 2
      McG: Discrim Score:      5.05
      GvR: Signal Score (-7.5): -3.14
      Possible site: 57
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program count: 0 value: 0.26 threshold: 0.0
10     PERIPHERAL Likelihood = 0.26      6
      modified ALOM score: -0.55

      *** Reasoning Step: 3

      ----- Final Results -----
15     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

20     33.2/56.1% over 281aa

      Bacillus subtilis
      EGAD|109228| hypothetical protein Insert characterized
      GP|2635109|emb|CA314605.1||Z99117 alternate gene name: yrdP Insert characterized
25     GP|1934657|gb|AA80908.1||U93876 hypothetical protein YrdP Insert characterized
      PIR|E69725|E69725 potassium uptake trkA - Insert characterized

      ORF01799(310 - 1128 of 1725)
      EGAD|109228| 82656(2 - 283 of 345) hypothetical protein { acillus subtilis}
30     GP|2635109|emb|CA 14605.1||Z99117 alternate gene name: yrdP { acillus subtilis}
      GP|1934657|gb|AA 80908.1||U93876 hypothetical protein YrdP { acillus subtilis}
      PIR|E69725|E69725 potassium uptake trkA - acillus subtilis
      %Match = 6.1
      %Identity = 33.2 %Similarity = 56.0
35     Matches = 77 Mismatches = 88 Conservative Sub.s = 53

      270      300      330      360      390      417      444      474
      YYC*LVKYFILHYFOQGEDMKHYQTIIIGAGAAGIGFGSAMQRLGLTNFLIIEKGH-IGESFL-RWFRTPQITPSFTT
40     | ||| ||| ||| | : : : | : : | | : | : | : | : :
      MYDTIVIGAGQAGISIGYYLKQ-SDQKPIILDKSHVEGSEWIKRDRYDSVLTRSMYSS
      10      20      30      40      50

      480      510      540      570      600      630      660      690
      -----NQGFPDIDNAVIPDTSFAPSFEKEHLGSGVEYARYLQLVAAHYNLPIONETSVLSIDIKRDLPLVIKTSKGDPS
45     ||| : : : | : : | : | : | : | : | : | : | : | :
      LPMGLHEGKEGHPFSPKNEIV-----AYLKKYVKKPEIPIQLRTEVSVLRIKIKYPLIKTRKEEYQ
      70      80      90      100      110

      720      750      822      852      882      912
      ADYLIMATGSPQNPNFIDIKGADLS----MHYGVQDNF-HIKSDNPFPIIIGNESACDALTHLVILGNQVELYTDTPGR
50     | : | : | : | : | : | : | : | : | : | : | : | :
      TKNLVIATGQPIVNPISIS-KDLSENINQLHSSQYKNSKQLAYGNLVVGAGNSGA-----
      130      140      150      160      170

      942      969      996      1026
      KESNFPDPSISLS-PLTKERLKHQI-DHKKEYYSISEGKKAI--EIKQIG-----
55     : : | : | : : : | : : | : | : : :
      -----QIAVELSKERVLYLACSNKLVYPLMIKGRSIFWFDKLGVLHASHTSVGKFIQKGGDFVGHGELKHAIK
      180      190      200      210      220      230      240

      1068      1098      1128      1158      1188      1218      1248
      -----KQYQVIFDDGSTAESFHKPLISLGLFWLCHLIDGIALPEYDKNQLPIVTRDEESTVNNCFIIGPSL
60     : : | : | : | : | : | : | : | : | : | :
      QKEIILKKRVIAAQKBIIFKDSSTLE-VNNITWATGFRNPLWNINKGVLDQGRITIHRRGVSFVEGLYFGLFWQHKR
      260      270      280      300      310      320

```

SEQ ID 8974 (GBS284) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 10; MW 42.7kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 9; MW 67.6kDa).

GBS284-GST was purified as shown in Figure 225, lane 7.

5 Example 2086

A DNA sequence (GBSx2201) was identified in *S. agalactiae* <SEQ ID 6457> which encodes the amino acid sequence <SEQ ID 6458>. This protein is predicted to be NrgA-like protein. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
  INTEGRAL    Likelihood = -11.73    Transmembrane    7 - 23 ( 1 - 31)
  INTEGRAL    Likelihood = -6.42    Transmembrane    86 - 102 ( 82 - 108)
  INTEGRAL    Likelihood = -6.42    Transmembrane    324 - 340 ( 318 - 342)
  INTEGRAL    Likelihood = -5.26    Transmembrane    210 - 226 ( 207 - 229)
  INTEGRAL    Likelihood = -5.10    Transmembrane    113 - 129 ( 112 - 133)
  INTEGRAL    Likelihood = -1.49    Transmembrane    246 - 262 ( 246 - 263)
  INTEGRAL    Likelihood = -1.17    Transmembrane    183 - 199 ( 183 - 199)
  INTEGRAL    Likelihood = -0.43    Transmembrane    41 - 57 ( 41 - 57)
  INTEGRAL    Likelihood = -0.00    Transmembrane    265 - 281 ( 265 - 282)

----- Final Results -----
      bacterial membrane --- Certainty=0.5692 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9997> which encodes amino acid sequence <SEQ ID 9998> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CB15668 GB:Z99122 ammonium transporter [Bacillus subtilis]
Identities = 105/378 (27%), Positives = 181/378 (47%), Gaps = 41/378 (10%)

Query: 3 VKKGLFVFLLLCIILSMIRGVAFFYFQSLH-QSLTSRIITYQFVLTVLLITTAIWPMAY 61
      ++ G VF+ C L+ WLM G+A +Y G+ ++ S ++ F ++ + + W + Y
Sbjct: 1 NMQMDITVFMFFCALLVWMTFGLALFYGGWVKSKVLTAMHSFS-SIAIVSTVWVLFQY 59

Query: 62 FLAFEGHFKTVPFQEQADGQI-----VNCFLQCLFALVAVVMLIGSIIDR 107
      LAF + + A K + +FQ+ FA+ ++ G+ +R
Sbjct: 60 TLAFAFONSIIGLEWAGLKGVGFDPKDYSDTIPHSLFMPQTFPAVLTTAIIISGAFAER 119

Query: 108 VQTKRLLLAVVSWFLPAYFVPLAVLIWNSEGVFAKGVLDPSGGMVHLCAGLSSYILAHV 167
      ++ LL V W LYTP+ A+ +W G ++G LDE+CG +VR-S+G+ ++LA V
Sbjct: 120 MRFGAFLFSLVMSLIVTTPVAHWVWG-GWIGLQALDFAGNVVHISBGVAGLVLAIV 178

Query: 168 IKG-----SEHQHNKVKNDSLPLMLITFGWFGNMGFUGEWNSQAMILLINTIPAIIG 222
      +KG + HN + FLG LI FWRGFG +G + A+ +NT A
Sbjct: 179 LGRKKDCTASSPHNLIYT---FLOGALTFQWGFVNGSALTLDGVAMVAFINTNTAAAA 235

Query: 223 GQANTLAANKWNGKERTGSLNGIITGLVTSFAGWGYLLWQLLAVITFASLPTFYFD 282
      G W L ++ ++G I CLV T G+ + + + +
Sbjct: 236 GIAGWILVEMIINKPTMLGAVSGATAGLVAITPAAGFVTPFASIIIGIGGAVCGVGF 295

Query: 283 YVAKAFADDDVSSFGMNGIGGLGSLGVGLFKLSIMP-----VQLAL 326
      + K F DD + +FG+ +CGG G+ GLF + + Q+A+
Sbjct: 296 SLICKFGYDDALDAPGLKIGGTWGGIATGLFATTSVNSAGADGLFYGDASLIWQIVAI 355

Query: 327 ATTILLSIIMTYIISKAI 344
      A T + I+T+I K +
Sbjct: 356 AATYVVFVITVTVIIRIV 373

```


[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2087

A DNA sequence (GBSx2202) was identified in *S.galactiae* <SEQ ID 6459> which encodes the amino acid sequence <SEQ ID 6460>. This protein is predicted to be dUTPase (dut). Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2731(Affirmative) < success
      bacterial membrane --- Certainty=0.0000(Not Clear) < success
      bacterial outside --- Certainty=0.0000(Not Clear) < success

```

A related GBS nucleic acid sequence <SEQ ID 9471> which encodes amino acid sequence <SEQ ID 9472> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP: CAA72644 GB: YL1901 dUTase [Lactococcus lactis]
Identities = 67/144 (46%), Positives = 90/144 (61%), Gaps = 8/144 (5%)

Query: 40  RGFLRSQFNSEKLLPKRETHAAGYDLKVAKKIVBPGETIPLVPIKTKAYQWGGELVLI 99
RGF+--+ +P+R+T H+AGYD+ ++ I+P EI +V TG+ + EVL L
Sbjct: 3    RGPK--KLGONATPIRATKSGSIVDIASBETTVQDEIKMSTGLVQLGDDDEVKLL 59

Query: 100 YDRSSNPGRKGIVLINSVIGDYGNNYQVNEHGIAPQMONITQAVILEEGRIVQAFV 159
YDRSSNP K+GI LINSVGI ID DYY + Ni+ V + +GRI+Q VF
Sbjct: 60 YDRSSNPGRKGTALINSVIGIDISDYVYQEFK----GLFMNISKEPVTISKQRIMQGVF 114

Query: 160 APFLADDDQDQVHMRTGGFGSGTG 183
+L DDD A+G RTGGFGSGTG+
Sbjct: 115 VKYLTIDDDNANGKRTGGFGSGTG 138

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6461> which encodes the amino acid sequence <SEQ ID 6462>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2519(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-2358-

Identities = 115/148 (77%), Positives = 125/148 (83%)

Query: 36 MSKVRGPELVSSQPSNKKLLPKRETHAAGYDLKVAKTVIRPGEITLVPTGIKAYMQDGE 95
 M+K+RGPELVSS P+N+LLPKRET HAAGYDL VA+ I PGEI LVPTG+KAYMQ GE
 Sbjct: 1 MTKIRGPELVSSPTNPULLPKRETHAAGYDLVARAVITAPGEIKLVPTGVKAYMQDGE 60

Query: 96 VLYLYDRSSNPRIKKGIIVLINSVGVWIDGYNNQVNEGHIFAGQMNTIDQAVILEEGERIV 155
 VLYLYDRSSNPRIKKGI+LINSVGVWID DYY N+ NRGHIFAGQMNTID V L GERIV
 Sbjct: 61 VLYLYDRSSNPRIKKGI+LINSVGVWIDYYNGEANEHGHIFAGQMNTIDPVTVAWGERIV 120

Query: 156 QAVFAPFLADDDQATGMRTGGFGSTGK 183
 Q VF PFL+AD DQA G RTGGFGSTG+
 Sbjct: 121 QGVFMPFLADGDQARGERTGGFGSTGQ 148

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2088

A DNA sequence (GBSx2203) was identified in *S. agalactiae* <SEQ ID 6463> which encodes the amino acid sequence <SEQ ID 6464>. This protein is predicted to be RadA homolog (radA). Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2628 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB11863 GB: Z99104 DNA repair protein homolog [Bacillus subtilis]
 Identities = 285/453 (62%), Positives = 358/453 (78%), Gaps = 4/453 (0%)

Query: 1 MAKKKSVFTCQEGYQSPKYLGRCPNCSAWSPVEVEVQEVONARVSLANGKSRPTKLK 60
 MAK KS F CQ GY+SPK++G+CP C AW++ VEE+ + N R + + K
 Sbjct: 1 MAKTKSKFIQCGYSPKXMGKCPGCGAWNTMVEEIKKAPANRRAPFHSVQTQKPS 60

Query: 61 DVSSINYS--RTKTMDPEFNRLVGGVVGSLVLIGQDPGIGKSTLLQLVSTQLA-NKG 116
 ++SI S R KT + EFNRLVGGVVG GSVLVLIGQDPGIGKSTLLQLVS QL+ +
 Sbjct: 61 PITSIETSEEPVKTQLGEPFNRLVGGVVGKSVLVLIGQDPGIGKSTLLQLVSAQLSGSN 120

Query: 117 TVLVYSGEESABQIKLRSERLGDIINEFVLYASTNMQSIRSEIEKIKPDLIIDSQITM 176
 +VLV+SGEES +Q KLR++RLG + ++ET+M+ I S I++ P F++DSIQT+
 Sbjct: 121 SVLYISGEESVQKTKLRDLRLGGINNPSLHVLSEITMEYISSAIGQMPSPFVVDYSIQTVY 180

Query: 177 SPEBVSQGVSVQVRETAELMQLAKTNINATPIVGHVTEKOTLAGPRLMEHMDVTLYLF 236
 +++S GSVSVQRE TAKLM++AKT I PIVGHVTKRGH+AGPR+LEHMDVTLYLF
 Sbjct: 181 QSDITSAPGSVSVQVRETAELMQLAKTKGIPIPVGHVTKESGIAGPRLLEHMDVTLYLF 240

Query: 237 EGERHHTFRILRAVKNRPGSTNEIGIFEMSGGLVEVLNPSQVLFLEERLDGATGSAIVVT 296
 EGERHHTFRILRAVKNRPGSTNEIGIFEM+ GL EVLNPS++FLER G+ GS+I +
 Sbjct: 241 EGERHHTFRILRAVKNRPGSTNEIGIFEMREGLTEVLNPSFLEERSAGSAGSSITAS 300

Query: 297 MEGTRPILAEVQALVTPTVGNKAKRTTGLDFNRVSLIMAVLEKRCGLLQNDAYLKSA 356
 MEGTRPIL E+QAL++PT FGN +R TG+D NRVS L+MAVLEKR GLLLQNDAYLK A
 Sbjct: 301 MEGTRPILVEIQALISPTSPGNPRMATGIDHNRSLLMAVLEKRVGLLLQNDAYLKVA 360

Query: 357 GGVKLDEPAIDLAVAVLASSYKEKPTNPQESPIGEIGLGEIRRVTRIEQRNEASKLG 416
 GGVKLDEPAIDL++IASS++ P NP + PIGE+GLTGE+RRV+RIER+ EA+KLG
 Sbjct: 361 GGVKLDEPAIDLAVISLASSYKPTNPADCFIPEVGLTGEVRRVRIERQVKEAALKG 420

Query: 417 FTKIYAPKNSLAGTEIPKGINDIVGVTVTSQVLK 449

-2359-

F ++ P +L G PKGI+VIGV V++ L+

Sbjct: 421 FKRMIIIPANLIDGWTKPKGIEVIGVANVAALR 453

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6465> which encodes the amino acid sequence <SEQ ID 6466>. Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2191 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 416/453 (91%), Positives = 441/453 (96%)

Query: 1 MAKKKSVFFQJQBCGYQSPKYLGRCPNCISAWSSFVKEVEVQEVKNARVSLNGEKS RPTKIK 60

MAKKK+ F QJQBCGYQSPKYLGRCPNCISAWSSFVKEVEV+EVKNARVSL GEKSRP KLK
Sbjct: 1 MAKKKATFIQJQBCGYQSPKYLGRCPNCISAWSSFVKEVEVQEVKNARVSLAGEKSRPVLKIK 60

Query: 61 DVSSINYSRTKTDMDRFNRVLGGGVVFGSLVLIGGDPGIGKSTLLQVSTQLANKGTVLY 120

DV +I+Y RT+TDM EFNRLVGGGVVFGSL+LIGGDPGIGKSTLLQVSTQLANKGTVLY
Sbjct: 61 DVENISYHRQTQTDMSDFNRVLGGGVVFGSLVLIGGDPGIGKSTLLQVSTQLANKGTVLY 120

Query: 121 VSGEESAQIKLRSERLGDIDNEFYLVAETNNQSRISSEIEKIKPDFLIIDSIGTINSPEV 180

VSGEESAQIKLRSERLGDIDNEFYLVAETNNQ+IR+EIE IKPDFLIIDSIGTINSF++
Sbjct: 121 VSGEESAQIKLRSERLGDIDNEFYLVAETNNQAIETENIKPDFLIIDSIGTINSFDI 180

Query: 181 SSVQGSVSQVREVTAEIQLAKTNNIATPIVGHVTKBGTLAGFMLEHMDVTVLYFEGER 240

+ VQGSVSQVREVTAEIQLAKTNNIATPIVGHVTKBGTLAGFMLEHMDVTVLYFEGER
Sbjct: 181 TGVQGSVSQVREVTAEIQLAKTNNIATPIVGHVTKBGTLAGFMLEHMDVTVLYFEGER 240

Query: 241 HHTFRLRAVKNRPGSTNEIGIFEMQSGGLVEVLNPSQVFLERLDGATGSAIVVTMEGT 300

HHTFRLRAVKNRPGSTNEIGIFEMQSGGLVEVLNPSQVFLERLDGATGSA+VVTMEG+
Sbjct: 241 HHTFRLRAVKNRPGSTNEIGIFEMQSGGLVEVLNPSQVFLERLDGATGSAIVVTMEGS 300

Query: 301 RPILAEVQALVTPTVFGNAKRITTTGLDFNRVSLIMAVLEKROGLLLQNDAYLKSAGGVK 360

RPILAEVQ-LVTPTVFGNA+RTTTGLDFNRVSLIMAVLEKROGLLLQNDAYLKSAGGVK
Sbjct: 301 RPILAEVQSLVTPTVFGNARRITTTGLDFNRVSLIMAVLEKROGLLLQNDAYLKSAGGVK 360

Query: 361 LDEFAIDLAVAVAIASSYKPKTPNQESFPIGEIGLTGEIRRVTRIBQRINEAKLGFYKI 420

LDEFAIDLAVAVAIASSYKPKTP+PQE+P+GEIGLTGEIRRVTRIBQRINEA+KLGFYK+
Sbjct: 361 LDEFAIDLAVAVAIASSYKPKTPSQEAFPIGEIGLTGEIRRVTRIBQRINEAARLGFYKY 420

Query: 421 YAFKNSLAGIEIPKGDIVGVTTVSQVLKAVFS 453

YAFKN+L GI+IP+GI+V+GVTTV QVL AVFS
Sbjct: 421 YAFKQLQSIDIPQGIIVGVTTVQGVNLAVFS 453

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2089

A DNA sequence (GBSx2204) was identified in *S.agalactiae* <SEQ ID 6467> which encodes the amino acid sequence <SEQ ID 6468>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.3488 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2360-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA97750 GB:273419 hypothetical protein Rv1284 [Mycobacterium tuberculosis]
 5 Identities = 69/162 (42%), Positives = 100/162 (61%), Gaps = 2/162 (1%)
 Query: 3 TYFDNPLKTNQAYADLHGTALHPLEPKTKVAIVTCMDSRLHVAQALGLGDAHLRNAG 62
 T D++L N YA LE+ P +AIV CMD+RL V + LG+ G+AR++RNAG
 10 Sbjet: 2 TVTDDYLANNVDYASGF+KGPLEMPSPKHLIAVACMDARLDVYRMGLIKEGEARVIRNAG 60
 Query: 63 GRVDDVRLSLVISQQQLGTRREIVVLHHTDCGQFTTNEAFAAQLQRLDGVMDHGHDF 122
 VEDDV+RSL ISQ+ LGTRRI++LHHTDCG TFT++ F +Q + G+
 Sbjet: 61 CVVTDVIRSLAISQRLGLTRRIILHHTDCGMLTFTDDPFKRAIQDETGIPTWSP-ES 119
 15 Query: 123 FNDIEBSVREDVAKLHASPLIPDDVVISGAIYVDVTGRMVEV 164
 + D E VR+ + + + P + + G ++DV TQ++ EV
 Sbjet: 120 YPDAVEDVRSRLRIEVPVFTKHTSLRGFVFDVATGKLEEV 161

There is also homology to SEQ ID 6470:

20 Identities = 126/164 (76%), Positives = 146/164 (88%)
 Query: 1 MTTTFDNFLKTNQAYADLHGTALHPLEPKTKVAIVTCMDSRLHVAQALGLGDAHLRN 60
 + +YF++F+ NQRY LHGTALHP+KPKTKVAIVTCMDSRLHVAQALGLGDAHLRN
 25 Sbjet: 1 LMSYFHFMAANQAYVAHGTALHPLEPKTKVAIVTCMDSRLHVAQALGLGDAHLRN 60
 Query: 61 AGGRVTDVRLSLVISQQQLGTRREIVVLHHTDCGQFTTNEAFAAQLQRLDGVMDHGHDF 120
 AGGRVT+D++RSLVISQQQ+GTRREIVVLHHTDCGQFTTNE FA + LGVD+ G DF
 Sbjet: 61 AGGRVTDVRLSLVISQQQ+GTRREIVVLHHTDCGQFTTNEGFAKIHHEHLGVDVSGQDF 120
 30 Query: 121 LPFDIEBSVREDVAKLHASPLIPDDVVISGAIYVDVTGRMVEV 164
 LPF D+E+SVRED+AK+ AS LI DDVVI+GA+YVDVTG+M +V
 Sbjet: 121 LPFDVQVEDSVRDKMAKIRASSLISDDVVINGAVYVDVTGKMTQV 164

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 35 vaccines or diagnostics.

Example 2090

A DNA sequence (GBSx2205) was identified in *S. agalactiae* <SEQ ID 6471> which encodes the amino acid sequence <SEQ ID 6472>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0536 (Affirmative) < succ>
 45 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9473> which encodes amino acid sequence <SEQ ID 9474> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAC73407 GB:AB000137 putative oxidoreductase [Klebscherichia coli K12]
 Identities = 199/438 (45%), Positives = 286/438 (64%)
 Query: 1 MKKYDVIVLFGFGKAGKTLLAKLATQGKSVAMVERDDKMYGGTCINIGCIPTKTLLVSASK 60
 M KY +++GFGKAGKTLLA G VA++E+ + MYGGTCINIGCIPTKT++ A +
 55 Sbjet: 10 MNKYGAVIIGFGKAGKTLLAVTLAKAGMRVALIEQSNMYGGTCINIGCIPTKTLLVNDQQ 69
 Query: 61 NHDPQEAATIRNEVTSRLRAKNTFAMLDNKDVIDVYNAKARFISNKVVELTGGADQKELTA 120
 + DF A+ +NEV + LR KNF L + +DV + +A FI+N + + E+

-2361-

Sbjct: 70 HTDFVRAIQKNEVVFNLKKNKPNHNLADMPNIDVIDQQAQFINNHSLSRVHRPEKNLEIHS 129
 Query: 121 DVIIINTGAKSVQLPIGLADSQHYVDSTAIQELAHFKRLGIIGGNGNLGLSFATLYSEL 180
 + I INTGA++V PIPG+ + VYDST + L LP LGI+GGG IG+SPA++++
 5 Sbjct: 130 EKIPIINTGACVVPPIPGITTTTGQYDSTGLLMLKELGHLGILGGGVIYGFSPASMANP 189
 Query: 181 GSKVTVIDSQSRIFAREEREELSBMAQCYLEEMGISPTLSADIKSVQNEDEVDVVISPRDEK 240
 GSKVT++++ S RE+ ++++ L + G+ L+A ++ + + + V + E +
 10 Sbjct: 190 GSKVTILAEASLFLHEDRDIADNTIATILRDQGVDTILNAHVERISHENQVQVHSRPAQ 249
 Query: 241 LSFDAVLVATGRKPNTEGJALENTDIKLTERGATAVDVEYQQTSVENIFAVGDMVGQPOFT 300
 L+ DA+L A+GR+P T L EN I + ERGAI VD+ T+ +NI+A+GDV GG QFT
 Sbjct: 250 LAVDALLIASGRQPATASLHPENAGIAVNERGAIVVDKRLHTTADNIWAMGDVTSGLQFT 309
 Query: 301 YISLDDSRIVLIAYLNCDDKDYSLKRGAVPTSTFINPPLATVGLDEKTAKEKQYQVKSNSL 360
 YISLDD RIV + L + S +R VP S F PPL+ VG+ B+ A+E G ++ +L
 15 Sbjct: 310 YISLDDYRIVRDELLSGSKKSTDDKKVVPYVFMTPPLSRVGMTEBQARESADIQVVIL 369
 Query: 361 LVSAMFRAVNNDLRGIFKVVVDTEINLILGARLPGASHELINIIYMMDNKPIPTYPQ 420
 V+A+PRA V ND RG+ K +VD +T +LGA L +SHE+INI+ M MD +PY+ +
 20 Sbjct: 370 PVAATPRARVMDTRGVLAKEIVNKTQRLGASLLCVDGHEMINIVQAVMDAGLPSILR 429
 Query: 421 KQIFTHPTMVENFNDLFW 438
 QIFTHP+M R+ NDLF+
 25 Sbjct: 430 DQIFTHPSMSLSLNDLFS 447

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2091

A DNA sequence (GBSx2206) was identified in *S.agalactiae* <SEQ ID 6473> which encodes the amino acid sequence <SEQ ID 6474>. This protein is predicted to be glutamyl-tRNA synthetase (gluX). Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2245 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9475> which encodes amino acid sequence <SEQ ID 9476> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10953> which encodes amino acid sequence <SEQ ID 10954> was also identified.

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC31971 GB:U49789 glutamyl-tRNA synthetase [Bacillus subtilis]
 Identities = 273/491 (55%), Positives = 353/491 (71%), Gaps = 19/491 (3%)
 Query: 20 LANKIRVRYAPSPFTGLLHIGNARTALFNLYLARIHGGDFVIRIEDTDKKHVEDGERSQL 79
 + N+RVRYPAPSPFTGLHIGNARTALFNLYL+AR+ G3 F+IR+EDTD+KR++E GB+SQL
 50 Sbjct: 1 MONEVVRVRYAPSPFTGLLHIGNARTALFNLYLPAENQGGKFIIRVEDTDKKKNIEGGGSQL 60
 Query: 80 ENLRWLGMDWDSESPET---HENYRQSERLELYQRYIDQLLAEGKAYKSVYTESLAAERE 136
 L+WLGA+WDSE + + YRQSER ++Y+ Y ++LL +G AYK Y TESEI ERE
 55 Sbjct: 61 NYLKWLGIDWDSEVDVGGSGPYRQSRNDIYKYVYEELEKGLAYKCYCTEERLEKERE 120
 Query: 137 RQELAGETPRYINEFGIMSTSEKEAYIAEREAAGIIPVIRLAVNESGIYKWMVMKGDIE 196
 Q GB PRY + +++ B+E +IAE G P++R V E + + D+VKG+I

-2362-

Sbjct: 121 EQIARGENPRYSGKGRDLTQEEQEKPIAE---GRKPSIRFRVPEGKVIAFNDIVKGEIS 176
 Query: 197 FEIGNIGEDWVIQKDGYPYTNFAVVDHDMQISHVIRGDDHIANTPQLMAYVYALGWE 256
 FE IG D+VI KKG PYNFAV ID+ M++HV+RG+DHI+NTPEQ+M+Y+A GN+
 5 Sbjct: 177 FESDGIG-DFVIVKKGDTPTYNFAVAIDYLMKMTVIRGEDHISNTPKQIMYIQAQMD 235
 Query: 257 APQPGHMTLILINSETGKKLSKRDNTLQFIEDYRKKGYMSAVNFIALGWNPGGEEI 316
 PQPGHMTLILIN E+ KKLKRD + +QIE Y++ GY+ EA+ENI LIGW+P GEE+
 10 Sbjct: 236 LPQPGHMTLIVN-ESRKKLSKRDSEIIQFIQYKELGYLPEALFNICGLGWSVGEEL 294
 Query: 317 PSREQLINLFDENRLSKSPAAPDQKKMWSNDYLKADPFESVFLCKPFLERAGRL--- 373
 F+EQ I +ED NRLSKSPA PD K+ M+N Y+K D+V L P L+ +AG++
 Sbjct: 295 PTKEQFIETFDVNRSLKSPAIDMHKLKWNQYVKKLLDQVVELTLPHLQKAGKVGTE 354
 Query: 374 -----TDHAKKLVELYQPKLSADEIVPLTDLFPADFPETAEKIVMAETVPTVLSAF 428
 + KL+ LY QL RIV LITLFF D E + K V+ E VP VLS F
 Sbjct: 355 LSAEBQEWKRLISLYHQLSYGAEIVELTDLFPDEIETVQEAQVLEEEQVPEVLSGF 414
 Query: 429 KEKLVSLDEEFTDTITPQIKAVOKETGIRKRLMPPIRAVSGEHHGPELPTITYLIG 488
 KL L EFT D I IKAVOKETG KGL LMPPIRAVAG+ HGPBLP +I L+G
 20 Sbjct: 415 AAKLEEL--EFTPTDNIKASIKAVOKETGHRGKGLMPPIRAVAGTQYHGPQSLIELIG 472
 Query: 489 KRKSQVGHIDRM 499
 KE ++Q + N+
 25 Sbjct: 473 KETAIQRLMNI 483

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6475> which encodes the amino acid sequence <SEQ ID 6476>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1966 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 434/481 (90%), Positives = 459/481 (95%)

40 Query: 20 LANKIRVRYAPSPGTLHIGMARTALFNLYARHGGDFVIRIEDTRKRKHVEDGERSQL 79
 ++ IRVRYAPSPGTLHIGMARTALFNLYAR HGG P-IRIEDTRKRKHVEDGERSQL
 Sbjct: 1 MSKPIRVRYAPSPGTLHIGMARTALFNLYARRHGCTFIIRIEDTRKRKHVEDGERSQL 60
 Query: 80 ENLKLWGLNDWDESPETHENYRQSERLALYQYIDQLLAGKAYKSVTTEELAAERERQE 139
 ENL+KLWGLNDWDESPETHENYRQSERL LYQ+YIDQLLAGKAYKSVTTEELAAERERQE
 45 Sbjct: 61 ENLKLWGLNDWDESPETHENYRQSERLALYQYIDQLLAGKAYKSVTTEELAAERERQE 120
 Query: 140 LAGETPRVINEPIGMSERKEKAVIAEREAAGIPTVRLAVNESGIYKWTDMVKGDIEFEG 199
 AGETPRVINEPIGMS EK YIAEREAAGIPTVRLAVNESGIYKWTDMVKGDIEFEG
 50 Sbjct: 121 AAGETPRVINEPIGMSADEKAKYIAEREAAGIPTVRLAVNESGIYKWTDMVKGDIEFEG 180
 Query: 200 SNIGGDWVIQKDGYPYTNFAVVDHDMQISHVIRGDDHIANTPQLMAYVYALGWEAPQ 259
 NIGGDWVIQKDGYPYTNFAVVDHDMQISHVIRGDDHIANTPQLMAYVYALGWEAPQ
 55 Sbjct: 181 GNIGGDWVIQKDGYPYTNFAVVDHDMQISHVIRGDDHIANTPQLMAYVYALGWEAPQ 240
 Query: 260 FGHMTLILINSETGKKLSKRDNTLQFIEDYRKKGYMSAVNFIALGWNPGGEEIFSR 319
 FGHMTLILINSETGKKLSKRDNTLQFIEDYRKKGYM RAVNFIALGWNPGGEEIFSR
 Sbjct: 241 FGHMTLILINSETGKKLSKRDNTLQFIEDYRKKGYMSAVNFIALGWNPGGEEIFSR 300
 Query: 320 EQLINLFDENRLSKSPAAPDQKKMWSNDYLKADPFESVFLCKPFLERAGRLTDKAKK 379
 EQLI LFDENRLSKSPAAPDQKKMWSN+YLK+ADFE+V+ALCKPFLERAGRLTDKAKK
 60 Sbjct: 301 EQLIALFDENRLSKSPAAPDQKKMWSN+YLK+ADFE+V+ALCKPFLERAGRLTDKAKK 360
 Query: 380 LVELYQPKLSADEIVPLTDLFPADFPETAEKIVMAETVPTVLSAFKRLVSLSDER 439
 LVELY+PQLKSADEI+PLTDLFP+DFPRLTAEKIVMA ETV TVL AFK KL ++SDE+

-2363-

Sbjct: 361 LVELYKPOLKSADEIITLPLDFSDPPELTRAKEVMAGETVSTVLQAFAKLRAMSEDED 420

Query: 440 PTDITIPPQIKAVQKSTGKGNLFPFIRIAVSGEMHGPELPTIYLLGKEKSVQGHIDNL 500
F + IFFQINAVQKSTGKGNLFPFIRIAVSGEMHGPELPTIYLLG++KS++HI NNL

5 Sbjct: 421 PKPEMIFPQIKAVQKSTGKGNLFPFIRIAVSGEMHGPELPTIYLLGRDKSIRHINNL 481

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2092

- 10 A DNA sequence (GBSx2207) was identified in *S. agalactiae* <SEQ ID 6477> which encodes the amino acid sequence <SEQ ID 6478>. This protein is predicted to be d-ribose-binding protein precursor, fragment (rbsB). Analysis of this protein sequence reveals the following:

Possible site: 24
>>> May be a lipoprotein

15 ----- Final Results -----

bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

20

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAH15613 GB:Z9122 ribose ABC transporter (ribose-binding protein) [Bacillus subtilis]

Identities = 143/301 (47%), Positives = 205/301 (67%), Gaps = 1/301 (0%)

25 Query: 14 MSIVLILGCGKTGLNNSGNTQNVTKKSAKDLKLGVSISTNNFFVPMKDGIDKYAS 73
+S++L L T K + K+ +G+S+ST NNP+PV++K GI+K A
Sbjct: 5 VSVILTLSLFLTACSLLEPPQWAKPNSONKKEFTIGLSVSTLNNPFFVSLKGIKREAK 64

30 Query: 74 NKKISIKVDAQDDAARQADDVONFISCNVDAILINPVDKAIPTAKSANNANIPVILM 133
+ + + + DAQ+D+++Q DV++ I Q VDA+LINP DS AI TA++SAN +FV+ +
Sbjct: 65 KRGMKVIIVDAQNDSKCTSDVEDLIQGVDALEINPTDSSAISTAVESANAVGVFVVTI 124

35 Query: 134 DRGSEGGKVLTVASDNVAAKGMAADYAVKLGKKAFAELSGVPGASATVDRGKGPHSV 193
DR +E GK V T VASDNV G+MAA + KLKG AK EL GVPASAT +RG GFH++
Sbjct: 125 DRBAEQKQVETLVASDNVKGEMAAAFIADLKGKAKVBLEGVPGASATRRSGSFINT 184

40 Query: 194 AKSKLDILSSGSANFDRAKALNTQNMIOQHHDVQIIFAQNDMALGAQAQVKSAGLQNV 253
A KL+++QSA+FDR K L +N++QSH D+Q +FA NDMAALGA +A+ S+G+++
Sbjct: 185 ADQKLQVVTQGSADPRTKGLTVMENLLQGHDPDQAVFAHNDMALGALAEINSSG-IDI 243

45 Query: 254 LIVIGDQPDADHAIKKGDISATIAQOPAKMGSLAIAQAIDYDKKKVEKETISPIYLVTK 314
L+G DG DA +IK +SAT+ACQP +G++A +AA D GKVK-K +P+ L T+
Sbjct: 244 LVIGPDKNDAIASIKDRKLSATVAQPELIGKLAT+ADDILHSGKVKQKTISAPLKEIQ 304

No corresponding DNA sequence was identified in *S. pyogenes*.

SEQ ID 6478 (GBS203) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 12; MW 36.8kDa).

GBS203-His was purified as shown in Figure 208, lane 8.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2364-

Example 2093

A DNA sequence (GBSx2208) was identified in *S. agalactiae* <SEQ ID 6479> which encodes the amino acid sequence <SEQ ID 6480>. This protein is predicted to be galactoside ABC transporter, permease, protein (rbsC). Analysis of this protein sequence reveals the following:

```

5      Possible site: 14
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -11.15    Transmembrane    63 - 79 ( 52 - 85)
      INTEGRAL    Likelihood = -3.66     Transmembrane    111 - 127 ( 110 - 128)
10     INTEGRAL    Likelihood = -2.71     Transmembrane    168 - 184 ( 168 - 188)
      INTEGRAL    Likelihood = -2.44     Transmembrane    189 - 205 ( 188 - 205)
      INTEGRAL    Likelihood = -0.80     Transmembrane    17 - 33 ( 17 - 33)

      ----- Final Results -----
15     bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9287> which encodes amino acid sequence <SEQ ID 9288> was also identified.

```

20    The protein has homology with the following sequences in the GENPEPT database.
      >GP:CAB15612 GB:299122 ribose ABC transporter (permease) [Bacillus subtilis]
      Identities = 144/211 (68%), Positives = 182/211 (86%), Gaps = 1/211 (0%)

25    Query: 1  M G M I N G L F I S Y G K I A P F I V L A T M T I F R G A T L V Y S N G N P I T A G L S D S F L Q F L Q G G Y I V G 60
      +GM+NGL I+ GK+APFI TLATMT+FRG TLVY++GNPIT GL ++ PQ G+GY +G
      Sbjct: 113 L G M I N G L I T K G N A P F I A T L A T M T V F R G L T L V Y D G N P I T - G L G T N Y G P Q M F G R G Y F L G 171

30    Query: 61  I P P F V I L N P L T F I I L Y L L E K T A P G K S V Y A L G G N E K A A Y I S G I K I N K V K I I Y T I S G I N A 120
      I P P I M L F+IL++LLEH PG+ YA+GGNEKAA ISGK+ +VK++IY++G++
      Sbjct: 172 I P V P A I T M V L A P V I L A V L L E K T P F G R R T A I G G N E K A A L I S G I K V T R K V K M I Y S L A G L L S 231

35    Query: 121 S I S G L I I T S R L S S A Q P T A G S Y E M D A I A A V L G G T S L S G G K R I I G T L I G A L I G V L N N G 180
      +++G I+TSRL SAQPTAG SYE+DAIAAVLQOTSLSGG+GRI+GTLIG LIIG LNNG
      Sbjct: 232 A L A G A I L T S R L S S A Q P T A G S Y E L D A I A A V L G G T S L S G G R G R I V G T L I G V L I G T L N N G 291

      Query: 181 L N I I G V S A P W Q V V K G I V I L A V L D R P K V A 211
      L N ++GVS+P+Q V V K G I V I L A V L D R K A K A
      Sbjct: 292 L N I L G V S S F Y Q L V V K G I V I L A V L D R K K S A 322

```

40 A related GBS gene <SEQ ID 8977> and protein <SEQ ID 8978> were also identified.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2094

45 A DNA sequence (GBSx2209) was identified in *S. agalactiae* <SEQ ID 6481> which encodes the amino acid sequence <SEQ ID 6482>. Analysis of this protein sequence reveals the following:

```

      Possible site: 35
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.12     Transmembrane    75 - 91 ( 74 - 91)
      INTEGRAL    Likelihood = -0.64     Transmembrane    96 - 112 ( 96 - 112)

50     ----- Final Results -----
      bacterial membrane --- Certainty=0.1447 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
55

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2095

A DNA sequence (GBSx2210) was identified in *S.agalactiae* <SEQ ID 6483> which encodes the amino acid sequence <SEQ ID 6484>. This protein is predicted to be ribose transport ATP-binding protein rbsA (rbsA). Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.00 Transmembrane 401 - 417 ( 401 - 417)

----- Final Results -----
bacterial membrane --- Certainty=0.1001 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15611 GB:299122 ribose ABC transporter (ATP-binding protein)
[Bacillus subtilis]
Identities = 297/493 (60%), Positives = 375/493 (75%), Gaps = 1/493 (0%)

Query: 1 MKIDMRNISKSGFINKVLEKIDLELSQGIHALMGENGAGKSTLMNLTGLFPASTGTIV 60
M+I+M+I K+PQ N+VL + +L G+HALMGENGAGKSTLMNLTGL A G I
Sbjct: 1 MQIEMKDIHKTPGKNQVLSGVSPQLMPGEVHALMGENGAGKSTLMNLTGLHKADKQIS 60

Query: 61 IDGEERTFSNPQEAERFGISFIHQEMTWPEMTVLENLFLGREIKTTPGLLAKIMRQKA 120
I+G E FSNP+EAER+GI+FIHQE+N WPEMTVLENLFG+EI + G+L + N+ A
Sbjct: 61 INGETYFSPNPKAEHQGLAFIHQELNTPMTVLENLFLGKBISSKLGVLQTRMKALA 120

Query: 121 LETFKRLGVTIPLDIPIGNLSVGQQQMIEIAKSLNQLSLVMDPEPTALTRETNLFR 180
E F +L V++ LD G SVGQQQMIEIAK+L+ +++MDEPTAALT+RE LF
Sbjct: 121 KEQFDKLSVLSLQQRAGECSVGQQQMIEIAKALMTNAKVIIMDEPTALTREISKLF 180

Query: 181 VIRGLKQEGVGVVYISHRMEEIFKIDTFVTVMRDGVVDITKETS/THSDSLVKKMVGRL 240
VI LK+ GV +VYISHRMEEIF I D +T+MRDG VDT S T+ DE+VKMVGRL+L
Sbjct: 181 VITALKKGNGSVIVYISHRMEEIFALCDRIITMRDGKTVDTINISSETDFDEVKMVGRL 240

Query: 241 HDYYPEKHSGEIGPVAFEVSNL-CDNPFEDVSFYVRKGEILGFSGLMGAGRTVEMRTIPGI 299
+ YP++ +G FEV N +FEDVSFYVR GEI+G SGLMGAGRT+MR +FG+
Sbjct: 241 TERYPKRTFSLGDKVPEVKNAISVKGSPEDVSFYVRSGEIVGSGLMGAGRT+EMR+LPGV 300

Query: 300 DKKKSGKVIDDQFRTITTTSPQAIKQGIGPLTENRDEBGLLDLPNKNMTLPSTKDFSK 359
D+ +G++ I ++ I P +A+K+G+GF+TENRDEBGL+LD +I++N+ LP+ FS
Sbjct: 301 DRIVTGEIWIAGKTAIKNPQBNVKKGIGFTENRDEBGLLDTSIRNIALPNLSSPSP 360

Query: 360 HGFPEKSTSTTFVQQLNRLYIKSGRPDLEVGNLGGNQKQVVLAKHIGIAPKVLILDEP 419
G D K FV LI RL IK+ P+ +LGGNQKQV+AKWIGI PKVLILDEP
Sbjct: 361 KGLIDHKRAEFVLLIKRITIKTASPETHARHLGGNQKQVVIKHWIGIPKVLILDEP 420

Query: 420 TRGVTVGAKRKIYQLMNLADRGVPIVMVSSDLPRLGVSDRIMVHGEISGLSRKXA 479
TRGVTVGAKRKIY LNNEL +RGV I+MVSS-LPEILG+SDRI+V+HGEISGL+ +EA
Sbjct: 421 TRGVTVGAKREIYTIANLTERGVVIMVSSSELPEILGMSDRI+V+HGEISGLHAREA 480

Query: 480 DQEKVMQLATGCK 492
QE++M LATG+
Sbjct: 481 TQERINTLATGGR 493

```

There is also homology to SEQ ID 4768.

-2366-

SEQ ID 6484 (GBS407d) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 2-4; MW 72kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 5 & 6; MW 47kDa).

GBS407d-His was purified as shown in Figure 235, lane 9-10.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2096

A DNA sequence (GBSx2211) was identified in *S. agalactiae* <SEQ ID 6485> which encodes the amino acid sequence <SEQ ID 6486>. This protein is predicted to be high affinity ribose transport protein rbsD (rbsD). Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15      bacterial cytoplasm --- Certainty=0.2673 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAB15610 GB:Z99122 ribose ABC transporter (membrane protein)
    [Bacillus subtilis]
    Identities = 74/131 (56%), Positives = 95/131 (72%), Gaps = 1/131 (0%)

25 Query: 1 MKKTGIINSHLAKLADLGHTDRCVIGDLGLPVPVPIKIDLSITSGIPSPQEVLDIYLE 60
      Sbjct: 1 MKKGIINSHLAK+ DLGHTD++ I D GLPVP+G+ KIDLSL GLP+PQ+ + E
      MKKGIINSHLAKLADLGHTDKIVADAGLPVDPGVLIKIDLSLKPGLPAQDTRAVLAE 60

Query: 61 NILVEKVIILAEIKRANFQDLRLAKLONSVSIEYVSHNHLQMTQDVKAVIRTGENTP 120
      + VEKVI A EIK +N + ++ L L + IEY+SH K +T+D KAVIRTGE TP
30 Sbjct: 61 EMAVEKVIAAAEIKASNDEN-AKFLNENLSEQRITFLSHREEFKLLTKDAKAVIRTGEFT 119

Query: 121 YSNIIQSGVI 131
      Y+N IILQ+GV+
35 Sbjct: 120 YANCIQAGVL 130
  
```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2097

- 40 A DNA sequence (GBSx2212) was identified in *S. agalactiae* <SEQ ID 6487> which encodes the amino acid sequence <SEQ ID 6488>. This protein is predicted to be ribokinase (rbsK). Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have an uncleavable N-term signal seq

45 ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
50
  
```


-2367-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15609 GB:Z99122 ribokinase [Bacillus subtilis]
Identities = 132/293 (45%), Positives = 177/293 (60%), Gaps = 4/293 (1%)

5 Query: 1 MSNIVIGSISMDLVMTNRILAKGEITVFGQRFPMVPOGKGANQAVAIKLSQRDNITI 60
M NI +IGS SMDLV+ +++ K GEIV G F VPOGKGANQAVA RL + + +
Sbjct: 1 MRNICVIGSCSMDLVVTSIDKRPKAGETVIGTSFQTVFGKGANQAVAAAILGAQ---VFM 57

10 Query: 61 LGAIGRDSFGPILLNLLKNHVTTDFVGITP-SSSGVAQITLYNNDRRIYCPGANGKVD 119
+G +G+D +G +L+NL NV TD++ + + SG A I L DN I+ GAN +
Sbjct: 58 VGVGVDDHYGTALLNNLKAUGRVTDYMEFVTHPSGTAHVLAEGDINSIVVVGKANDDIT 117

15 Query: 120 TKKMSQENSIIEKADLVVLQNEIPHQANCKIANFCKRHSIKLLYNPAPSRRETDIEMLDKV 179
I++ D+V++Q KIP + ++ +C H I ++ NPAP+R E +D
Sbjct: 118 PAYALNALEQIEKIVKVMVLTQQEIPSEETVDEVCYCHSHDIPILINPAPARLKGQETIDHA 177

20 Query: 180 DYFTPNHEHECOELFPNQKLEEDILATYPEKILIVLGATKGAIYSIGKESHLIPAEITKAVDT 239
Y TPNEHE LFP + + LA YP KL +T G +G YS G + LIP+ + VDT
Sbjct: 178 TYLTPNEHEASLIPBELITSEALALYPAKLEITBKGQGVRYSAGSKEVLIPSPFVPEVDT 237

25 Query: 240 TGAGDTFNGAFGYAISKKFKIAKALRFATLAHLVQKFGAQQGMPTIKEMED 292
TGAGDTFN AF A+++ I ALRFA AA LSV FGAQQGMPT E+H+
Sbjct: 238 TGAGDTFNAAFVALABEGKDIEAALRFANRAASLSVCSFGAQQGMPTIRNEVEE 290

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2098

30 A DNA sequence (GBSx2213) was identified in *S.agalactiae* <SEQ ID 6489> which encodes the amino acid sequence <SEQ ID 6490>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2272 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9477> which encodes amino acid sequence <SEQ ID 9478> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15608 GB:Z99122 transcriptional regulator (LacI family)
[Bacillus subtilis]
Identities = 141/327 (43%), Positives = 204/327 (62%), Gaps = 4/327 (1%)

45 Query: 13 MSTIRQVAEKAGVSTSTVSRIYSONGYVVSQKASQKIEQAIRESLHYVFNPLAQSLKTKKIQ 72
M+Ti+ VA AGVS +TVSR ++ NGVY ++ ++ A+ +L+Y PN +A+SL +++++
Sbjct: 1 MATIKDVAGAGVGSVATVSRNLNDNGYVIEHETRTRVIAAAMKLNYTPNEVARSLYKRSR 60

50 Query: 73 LVGLLLPDISNPPFPRIARGVEEFLKQGYRVMGLGNINNKSHLKEEYLVNLQSNAGII 132
L+GLLLPDI+NDFFFP+LARG E+ L +GYR++ GN++ + E EYL Q++ AGII
Sbjct: 61 LVGLLLPDIITNPPFPQLARGAEDELNREGYRLLIPGNSDEELKKRLRYLTPTKQKNHVAGII 120

55 Query: 133 --TTHDFTKNIHEIDIPVVVVDRVNOETQYGVPSDNKSGKLAQAQIWTGATNILLIRG 190
T + + + + PVV +DR E V SD G KLAQAQI + I L+RG
Sbjct: 121 AATNYPDLIEYSGKNYPVVVFLDR-TLKGAPSVSSDGYTGKLAQAQIHTKXSQRITLLRG 179

Query: 191 FLDKADNINRQPGSQNYLLNKGACFAIEDSASFDFARIQTAKTLLDHPFDIDSIAPS 250

-2368-

P RF G+ L F + ++ASF + Q AK L +P D +IA +
 Sbjct: 180 PA-HLP TADQDFNGALRLKQAEVD PQVISTASFSDIDQASMAKRLFASTPATDGVIA SN 238

Query: 251 DIHAIAYLHELINRGKRIPEDVQIIIGYDDILMSQFIYPSLSTIHQSSYIMQQAELIFK 310
 DI A A LHE L ROK +PED+QIIGYDDI S ++P LSTI Q +Y MG++AA+L+
 Sbjct: 239 DIQAAVLHELALRRGNVPBDIQIIGYDDIQSGLLFPPLSTIKQPAYDMGKEAAKLLG 298

Query: 311 ITNQLPITNKRIKLPHVYVERETLRK 337
 I + P+ I++PV Y+ R+T R++
 Sbjct: 299 IIRKQLABTAIQMPVTYIGRKTTRKE 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6491> which encodes the amino acid sequence <SEQ ID 6492>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1657 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 232/328 (70%), Positives = 274/328 (82%)

Query: 10 GVSMTIRQVAKGAVSTSTVRYISQNGYVSQKASQIBQAIRELHYVPNPLAQSLATK 69
 G +M TI+QVAB+AGVS STVRYISQ GYVS A KI+ AI +LHY FN LAQSLATK
 Sbjct: 14 GRAMVTIKQVABEAGVRSSTVRYISQNGYVSDARHKIKAAIAKLHYTPNVLQSLATK 73

Query: 70 KNLVGLLLPDISNPPFFRLARGVEEFLKEQZGRVNLGNVYNNKSHLEBYLNVLLQSNAA 129
 KNLVGLLLPDISNPPFFRLARG EE+LKE+GKRVNLGN ++ LEEBY++VLLQSNAA
 Sbjct: 74 KNLVGLLLPDISNPPFFRLARGAEYLNKRGYRVNLGNISDSALEEYVHVLLQSNAA 133

Query: 130 GIITTHDFIKHPEIDIPVVVVDKRVNQBETQYGVFSIDNKBGKLAQAQIWTAGATNILLIR 189
 GIITTHDFIK +P + IPVVVVDKRV+QETQYGVFSIDN+ GG LAAQ +M AGA +LLIR
 Sbjct: 134 GIITTHDFIKRYPTLAIPVVVVDKRVQETQYGVFSIDNAGGLAAQTWQAQAEVLLIR 193

Query: 190 GPLDKADNLNORFQSGQNTLLNKGACFAIEDSASFDPAEIQIBAKTLLDHHDDIDSIAP 249
 GPLD A+N+N+RF+ S +YL + + DS +FDF IQ+BA L +P IDSIAP
 Sbjct: 194 GPLDNABNINERFEASFYTLQKQVMTVMVCDSENFDFESTIQLEASYNLEKCYPTIDSIAP 253

Query: 250 SDIHAIAYLHELINRGKRIPEDVQIIIGYDDILMSQFIYPSLSTIHQSSYIMQQAELIF 309
 SDIHAIAY+HE+ ++GK+IP+DVQIIIGYDDILMSQFIYPSLSTIHQSSY+MG+ AAKL++
 Sbjct: 254 SDIHAIAYIHELMSGKIKIPQVQIIIGYDDILMSQFIYPSLSTIHQSSYLMGRYAEVLVY 313

Query: 310 KITNQLPITNKRIKLPHVYVERETLRK 337
 I +QL + RIKLPHVYVERET+R++
 Sbjct: 314 TIASQLTVKANRIKLPHVYVERETLRK 341

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2099

A DNA sequence (GBSx2214) was identified in *S.agalactiae* <SEQ ID 6493> which encodes the amino acid sequence <SEQ ID 6494>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -13.80	Transmembrane	27 - 43 (24 - 51)
INTEGRAL	Likelihood = -10.61	Transmembrane	337 - 353 (329 - 362)
INTEGRAL	Likelihood = -9.18	Transmembrane	257 - 273 (249 - 276)
INTEGRAL	Likelihood = -8.92	Transmembrane	302 - 318 (291 - 326)

-2369-

----- Final Results -----

bacterial membrane --- Certainty=0.6519 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 8979> which encodes amino acid sequence <SEQ ID 8980> was also identified. Analysis of this protein sequence reveals the following:

10 Lipop Possible site: -1 Crend: 6
 SRCFLG: 0
 MoG: Length of UR: 4
 Peak Value of UR: 3.20
 Net Charge of CR: 1
 MoG: Discrim Score: 6.06
 GvH: Signal Score (-7.5): 0.0500002
 15 Possible site: 46
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 47
 ALON program count: 3 value: -10.61 threshold: 0.0
 INTEGRAL Likelihood = -10.61 Transmembrane 325 - 342 (318 - 348)
 20 INTEGRAL Likelihood = -9.18 Transmembrane 246 - 262 (238 - 265)
 INTEGRAL Likelihood = -8.92 Transmembrane 291 - 307 (280 - 315)
 PERIPHERAL Likelihood = 4.98 152
 modified ALON score: 2.62
 icml HYPID: 7 CFP: 0.525
 25 *** Reasoning Step: 3
 ----- Final Results -----

30 bacterial membrane --- Certainty=0.5246 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAF12525 GB:AE001863 hypothetical protein [Deinococcus radiodurans]
 Identities = 103/352 (29%), Positives = 191/352 (54%), Gaps = 9/352 (2%)
 Query: 15 ANKSLTFYKKYLLIELLIIVMFVFLSGLANGLGRAVSAIENNPATQYIILNGBAQ 74
 A +E+ K + LLI ++ ++ FMV L+GL GL R ++ + + PAQ++ + A+
 40 Sbjct: 4 ALRELQHQKLRSLILGGIVALIAPMFVFLITGLTGLRSDSASILLITPAQS FVITKEADG 63
 Query: 75 VITSSVLITKDYTLNLSNLKDDSTTLNQRSSILTRQHEKKDISYFAIDKDSFAPITLS 134
 V+ S L+ + +++L + ++ ++ +K++ +D F+AP +S
 45 Sbjct: 64 VLNRSLFPEQ---VSLQQQNEAFAAQTFFVPSHODQLSGVLLGVDPGRFLAPDVS 120
 Query: 135 ECKQLTSYKKAIIILNLSKASIKLGDIVDKSSSISLTVVGFVHNSYKHGHPVAFIDKD 194
 EG+ L A++ ++SL+ +G+K+GD + K S L V GP ++ H P ++
 50 Sbjct: 121 EGQTLRVAGGAVV-DESLREDGVKQDVLITKPSGDQLKRVSGFTRARLNHQFGWIVSLA 179
 Query: 195 IYTEINKKINPOYQFLPQALVKNKDKSISHLP-TQLEAVSKKDVIGHIPYISABQSTIAN 253
 + +K+NP+ A++ + +L L++ +Q +EGY EQ +L M
 55 Sbjct: 180 RW----QLKLNRMHGTVAVALPAPAPQVNLGGRLSVTRHAQTLQVLQYKREKQSLITM 235
 Query: 254 ILNVLVVASAGILGVFFPIITLQKRHEPSVMKAGCTKMSIALPQLSGVILALFGIIV 313
 I L+ +A +L FFY++TLCK +F ++KAIG +A ++Q++IL L + +
 60 Sbjct: 236 IQVFLIARAAPFLATFFVFMVITLQRTAQFLKAKAGASNETLAGSVVAGMLITLLAVATA 295
 Query: 314 DGLAVALSIVLPQMPFVIMQNTILVSFVFLVAMISSALSIVKAKIDPV 365
 + + + +LFA MFF + NI S + LV+A ++S LS+ +VAK+DP+
 Sbjct: 296 AAVTLGVQLLPAGMPFILTAAIASASGILLVVAALSLASVRRVMVDPL 347

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6495> which encodes the amino acid sequence <SEQ ID 6496>. Analysis of this protein sequence reveals the following:

Possible site: 58

-2370-

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Like:lihood = -12.31	Transmembrane	246 - 262 (233 - 270)
INTEGRAL	Like:lihood = -8.49	Transmembrane	327 - 343 (321 - 351)
INTEGRAL	Like:lihood = -1.01	Transmembrane	301 - 317 (301 - 317)

----- Final Results -----

bacterial membrane	---	Certainty=0.5925(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

>GF:AAF12525 GR:AB001863 hypothetical protein [Deinococcus radiodurans]
 Identities = 101/360 (28%), Positives = 175/360 (48%), Gaps = 11/360 (3%)

Query: 1 MFLALNMKQSKLRGLIAGLLCLVAYLMFFLSGLAPGLMQENRKAQVMDLWADSVLLAKD 60
 M+LAL E++ KLR LI G++ L+A+++F L+GL GL +++ S + AS + K+
 Sbjct: 1 MYLARRELQHQKLRSLILGGIVLAIAMVFMILTGLTRGLSRDSASLLDLTAQSFVYTKK 60

Query: 61 ADATITLSQVSRACENQITADKVAPLAQLNTVAMSVKNPKDADKVKVSLPGIDNSFIR 120
 AD L S +S Q ++ D A T K V L G+D P+ P
 Sbjct: 61 ADGVLNRSFLSPQVSAQDQEDNEDAAAFATFVSPFSGDKQLSGV---LLGVDPRGFLAP 117

Query: 121 NIVKRLFKTNKEVVDQSLAKEEAFPAIGKDFYTSSSQALITVGYTQARFSPVAPVVM 180
 ++ +G+ + V+D+SL +E+ +G S L G+T++AR + P +Y+
 Sbjct: 118 DVSEKQLTKVAGGAVVDBSL-REDGVKVGDLVTLKPSGDQLKVGSPTRSRINMQHGMV 176

Query: 181 NLEAFETLKYGEPLPKDKQVNAFITKGS--LTDPYKKDFKLDIKFTIKPLFGYSAQL 238
 +L ++ L P+ VNA + + D + + LFGY +
 Sbjct: 177 SLARWQKLN-----PRMGITGVNAVALPAAPQVNLGGADLSVTNRQAQLTLQFGYKEBQG 231

Query: 239 TGFGMISPLVIISAIHIGIMYILTIQKAPIGIMKAQGISNKTITTAVMQTFFLSFLG 298
 + + FL+ ++A + F Y+T+QK FG++KA G SN+T+ +V+ Q L+ L
 Sbjct: 232 SLTMIQVFLTAVAAPVLATFFVYMLTKTAQGLLKAIGASNTLAGSVAGMILITLLA 291

Query: 299 SGLGLLGTWLTSLLLTPTVPFQSNWFLYLAIFVSMICFALLGLEFVSFNIIRIDPLKAI 358
 + T LLF +FF + ++ A L +L SV + +DPL A+G
 Sbjct: 292 VAIDAAVTLGMVOLLFAGMPFHLTAANTASAGLLLVAAALASLVRRAVAKVDPLIALG 351

An alignment of the GAS and GBS proteins is shown below.

Identities = 96/356 (26%), Positives = 178/356 (49%), Gaps = 4/356 (1%)

Query: 15 AWKELTFYKKCYLLIELLIIVMMFVVFSLGSLANGRAVSPAETNNPAQTYILNEGAEQ 74
 A E+ K +Y LI L+ ++ ++ FLSGLA GL + ++A+ A + +L +A+
 Sbjct: 4 ALNEMKQSKLRGLIAGLLCLVAYLMFFLSGLAPGLMQENRKAQVMDLWADSVLLAKDADA 63

Query: 75 VITSSVLITKQDILNSINLKDSTLTINQRSSLTRQGEKKIDISYPAIDKOSFMAPTSL 134
 +T S ++ + + + + LN S+ K+ +S F ID +SF+ P +
 Sbjct: 64 TLTLSQVSRACENQITADKVAPLAQLNTVAMSVKNPKDADKVKVSLPGIDNSFIRPIIV 123

Query: 135 EGGKQLTSYKKAIIINDSLKAEGIKLSDKVIDKSSSILITVVGPFVNSMYGHGFAVIDKD 194
 +G+ + K+ ++ KE +G SSS +LT+VG+ N+ + PV +++ +
 Sbjct: 124 KGRLFKTNKEVVDQSLAKEEAFPAIGKDFYTSSSQALITVGYTQARFSPVAPVVMNLE 183

Query: 195 IYTEIN-KKINPQYQFLQALVMKNKDSISHLPTQ-LEAVSKKVIQIHPIGYSAGQSTLN 252
 + + + + P+ ++ A + K S++ P + + + K I +PGYSA+ T
 Sbjct: 184 AFETLKYGEPLPKDKCVNAFITKGS--SLDTPKKDFKLDIKFTIKPLFGYSAQLLFG 241

Query: 253 MILWLVVASAGILGVFFYYIITLQKRHSFVMAIGTKMSEIALPQLSQVILLALFGIIV 312
 ++ LV+ SA I+G+F YI+T+QK F +MKA G I L Q L+ G+ +
 Sbjct: 242 FMISPLVLIISAIHIGIMYILTIQKAPIGIMKAQGISNKTITTAVMQTFFLSFLGSGGL 301

Query: 313 GDGLAVALSYVLEAPMPFVINWQNIILVSVFVLVAMISSALSIVKAKIDPVEVI 368
 G S +L+P +PF NW ++ + +A++ S+ + +IDP++ I
 Sbjct: 302 GLLGTLWLTSLLLTPTVPFQSNWFLYLAIFVSMICFALLGLEFVSFNIIRIDPLKAI 357

-2371-

SEQ ID 8980 (GBS239) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 13; MW 64kDa).

GBS239-GST was purified as shown in Figure 227, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2100

A DNA sequence (GBSx2215) was identified in *S. agalactiae* <SEQ ID 6497> which encodes the amino acid sequence <SEQ ID 6498>. This protein is predicted to be heterocyst maturation protein (devA) (b0879). Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1751 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA05977 GB:AJ003195 ATP-binding subunit [Anabaena variabilis]
Identities = 87/225 (38%), Positives = 146/225 (64%), Gaps = 1/225 (0%)

Query: 3 AILELKHISGHYPGDELLSILNLDLSVSGAEFVAIIGPSGGKSTLLSIAGLLLGADQ 62
      A++ +K ++ +Y G IL +++L + GE V + GPSGSGK+TLLS+ G L +
Sbjct: 5 AVIAIKSLMHYYGKALKRQILFDINLIYPGRIIVITGPGSGGKSTLLSIAGLLRVSQ 64

Query: 63 GSLVYNHENVITLDSQRQTQLRREALGFIPQSHQLLPYLTIQEQQEARFAKHDKTS 122
      G+L ++ SQ + Q+RR ++G+IQ+H LL +LT ++ +Q ++H ++ +
Sbjct: 65 GNLQFLGVELSGASQNKLVQIRR-SIGYIFQAHNLLGLFTRQNVMQAVELNEHISQERA 123

Query: 123 LSEINKILSDLGIBCCAHKYPNQLSGGQKQRAAARAFINHPKVLADSEPTASLDEERGR 162
      + + +L +G+E YP+ LSGGQKQK AINRA +N+P ++LADSEPTA+LD++ GR
Sbjct: 124 IAKAEAMLKAVGLNRYDYPNLSGGQKQKRAAARALVNNPPLVLADSEPTAALDKQSGR 163

Query: 183 QVTEILIRQEVKSHNTAIAIMVTHDERVLVDVTVRKLKGLVKEN 227
      V E+++ K T+ ++VTHD R+LD+ D + ++DG L +++
Sbjct: 184 DVVEIMQRLAKDQGTBILLVTHDNRILDIADRIVEMEDGILARDG 228

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6499> which encodes the amino acid sequence <SEQ ID 6500>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4161 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 103/224 (45%), Positives = 149/224 (65%), Gaps = 4/224 (1%)

Query: 3 AILELKHISGHYPGDELLSILNLDLSVSGAEFVAIIGPSGGKSTLLSIAGLLLGADQ 62
      ++L K ++K + DG ++ L D S+ AGEFVAI+GPSGGKST L+TAG L
Sbjct: 3 SVLTFKQVTKTFPGDGHRIINALKATDPSIRAGEFVAIIGPSGGKSTFLTIAGGLTPSS 62

Query: 63 GSLVYNHENVITLDSQRQTQLRREALGFIPQSHQLLPYLTIQEQQEARFAKHDKTS 122

```

-2372-

```

      G L +- + T Ls+++R++LR +++GFI Q+ L+P+ T+Q+QL+ H
Sbjct: 63 GQLIIDGTDTYHLSEKRRSLRFRKSGVFLQASNLIPFSTVQQQLR----LVNHLTGSKE 118

Query: 123 LERINKLLSLLGIEQCAHKYPNQLSGGCKQRAIARAFINHPKVILADEPTASLDEERGR 182
      + N+L DLGI H+ P +LSGG++QRAIARA + P +ILADEPTASID E+
Sbjct: 119 KAKANQLFDLLSTGLKHQLPQLSGGCKQRAIARALYHDPALILADEPTASLDTKAY 178

Query: 183 QVTELRQSVKSHNYFAIMVTHDERVLVDVTVYRLKDGKLVKE 226
      +V +L+ +E K N A IMVTHD+R+L D VYR++DG+L +E
Sbjct: 179 EVVKILAKESKERNKALIMVTHDDRMKLYCDRVYRMDGGLQGE 222

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2101

- 15 A DNA sequence (GBSx2216) was identified in *S. agalactiae* <SEQ ID 6501> which encodes the amino acid sequence <SEQ ID 6502>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2645(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

- 25 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB64972 GB:AJ012050 VicR protein [Enterococcus faecalis]
Identities = 86/229 (37%), Positives = 132/229 (57%), Gaps = 10/229 (4%)

Query: 3 KILVVEDNIVQCKIITFKLTQEGYQFITASNGQALNCLDTEEVQLIITIMMNMMDGYQ 62
      KILVV+D +I+ L +EGY+ TA +G+RAL ++ E LII D+M+P MDG +
Sbjct: 52 KILVVDDEKPISEIVKYNLVKGEYVFTAYDGEALKEVVEVPDLIILDLMLPFMDGLE 111

Query: 63 LIQELRSAAYNVPIIWM TAKSQMEDMTKGFGLGADDYVMKPVQLQELALRIKALLR-- 119
      + +R+R +++PII++TAK D G LGADDY+ KP +EL R+KA LHR
Sbjct: 112 VAREVRK-THMPIMVTAKDSEIDKVLGELGADDYVTKPFPSNRELVARVANLRGAT 170

Query: 120 ----ANIVAQHLIIGNICLNEDELSLKYFBQSIIFPCKEPRVLPHLLSYNRIPTRELEL 175
      A + Q +L IG+ ++ D + ++I +EF +L++L + ++ TR L
Sbjct: 171 NAKAEAVTQSELTIGDLTHPDATMVSERGEKIELTHREFELLYLAKHIGQNTREHL 230

Query: 176 LDSINGDTDLDERVVDACINKIRKVEHLDPFK--IRTVRGSGYRAKN 222
      L ++WG D D R VD + ++R K+E P + T RGVGY +N
Sbjct: 231 LQTVWGYDYPGDVRYVDVTVRRLREKIEDSPSHPTYLVTIRRGVGYLRN 279

```

- 45 There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2102

- 50 A DNA sequence (GBSx2217) was identified in *S. agalactiae* <SEQ ID 6503> which encodes the amino acid sequence <SEQ ID 6504>. This protein is predicted to be sensor protein. Analysis of this protein sequence reveals the following:

```

Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -8.97 Transmembrane 53 - 69 ( 47 - 77)

```

-2373-

----- Final Results -----

bacterial membrane --- Certainty=0.4588 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC62214 GB: AF049873 sensor protein [Lactococcus lactis]
 Identities = 97/307 (31%), Positives = 169/307 (54%), Gaps = 16/307 (5%)

Query: 57 SALAVVFLSLVSIASISGMWYGSYELTKPILDISHIVENADGDFSGHIYRNSNRKRSYFY 116
 + LAV+ +L++ +S++Y + +T+P+L I +A GD + N+
 Sbjct: 170 AVLAVI--TLIVTAFSIFITYITRTVTRPLLKIKLSTDKIAQGLSIQLNVNR----- 219

Query: 117 NELDELSEINQMIVLSLHMDENRKDFITNVSHLKTPLAANIVIELQLDPELDEETQS 176
 +EL EL++SI + L M R +F++V+HEL+TP+ + ++ E ++
 Sbjct: 220 DELGLAKSIEDLAKLDPMKRENEFLSSVAHELRTPLTFIKGYADIANRSTTSLEOKT 279

Query: 177 ELGLAVKTESLRLLTRICDTMLQMSKVDNQETIGELSSVRVDEQIRQAMISLTERWQAKRI 236
 + L +++ ES LT+L + + +++++ E V + E I + + + + KRI
 Sbjct: 280 QYLRIRESRHLTLQMEDLMQLAQESNGFKVEKHQVLQELINEVSVKSGVFSEKRI 339

Query: 237 NFQLDSKPYTVVYNSDLIM--QVWINLLNAIKYSEDIVLSVRMEETNNHYLRVVISDK 294
 NF L S Y+N D + QV +NLL NA KYS D D + + + + + ISDK
 Sbjct: 340 NF-LISGEGNFYANIDFMRIEQVLVNLNAYKSADESDIKLAFIPEKENF-KIVISDK 397

Query: 295 GRGISQYDVQHIFDKFYQADQSHNQ--GNGLGLAIVKRIIVLCKGRISVSQLEIGTF 352
 G GI + D + IP++FY+ D+S + G GLGLAIV+ I+ G+I V S GT F
 Sbjct: 398 GSGIPQDLFYFERFYRVDSKIRTTGGVGLGLAIVQDIVKKNKGIIVESIQNGITP 457

Query: 353 CVELPLS 359
 +ELP S
 Sbjct: 458 IIEPLYS 464

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8981> and protein <SEQ ID 8982> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 MG: Discrim Score: 4.84
 GVH: Signal Score (-7.5): 0.179999
 Possible site: 35
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 1 value: -8.97 threshold: 0.0
 INTEGRAL Likelihood = -8.97 Transmembrane 50 - 66 (47 - 77)
 PERIPHERAL Likelihood = 1.27 324
 modified ALOM score: 2.29

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4588 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

31.9/57.3% over 293aa

Lactococcus lactis

GP|3687664| sensor protein Insert characterized

ORP01881(478 - 1377 of 1677)

GP|3687664|gb|AAC62214.1|AF049873(171 - 464 of 464) sensor protein [Lactococcus lactis]
 %Match = 12.9

-2374-

```

%Identity = 31.9 %Similarity = 57.3
Matches = 94 Mismatches = 121 Conservative Sub.s = 75

5 339 369 399 429 459 489 519 549
  MKLRRFPFRLPYFTMLFVSLASILLVAIVPFPQGVITTHVLQVSAVLAVFLSVIASISMWVGSYHLTKP
  | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
  EKNKKSLSLHFHLGDKYIYKRSIQSNQKIVGVSVMPLSTRPIQKVMFNPFGI FAVLA VITLIVAFPSIFPITRTVIRP
  130 140 150 160 170 180 190

10 579 609 639 669 699 729 759 789
  ILDISHIVSNVADGDFGHYRNSNRKSYEYNELOELSSINOMIVLSIMDMKDFITNVSHELKTPIAAVANIVE
  : | : : | : | : : : : : | : | : | : : | : | : : | : | : | : | : | : | : |
  LLKIKLGTDKLAQGLSIQLNVTE-----DELSELAKSIDLAEKLDPMKHERNELSSVAHELKTPITLPIKGVAD
  210 220 230 240 250 260

15 819 849 879 909 939 969 999 1029
  LLQDPPELDEETSGELLGLVKTSLRLTRLCDTLMQMSKVDINGETIGELSSVRVDEQIRQAMISLTERWQAIRNFOLDK
  : : : : | : : : | : | : : : : : | : | : : : : : | : : : : : | : | : | : |
  IANRSTTSLEKTKQLRIEESRHLTLMEDLMNLAQLEENGPKVEKHQVLIQELINEVSVKSGVSEKRIKF-LISG
  280 290 300 310 320 330 340

20 1059 1083 1113 1143 1173 1203 1233
  PFTVYSNGLL--MQVWIMLLDAIKYSEIDVLSVRMESTNNHFLRVISDKGRGISQYDVQHIFDKPYQADQSHNQQ-
  : : : : | : : | : | : | : : : : : : : : | : : | : : | : : | : : | : : | : : |
  ENFPIANIDFRIEGLVNLMLNAYKSADESDIKLAFIPEKNF-KIVISDNGSGIPEQDLPYIFERFVRDSEKRTTT
  360 370 380 390 400 410 420

25 1287 1317 1347 1377 1407 1437 1467 1497
  -GGGLGLAIVKRIIVLCKGRISVSSQLEIGTFPCVLEPLS*LFKTTITAMQLLFLFRNKTYKNRKL*KYLITNIASV*
  | | | | | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
  GGVLGLAIVQDIAVKNGKILIVSISQNGSTTFIILPLPS
  440 450 460

```

SEQ ID 8982 (GBS170d) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 181 (lane 4; MW 35kDa) and in Figure 123 (lane 5-7; MW 35kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 2-4; MW 60kDa) and in Figure 184 (lane 3; MW 60kDa). Purified GBS170d-GST is shown in Figure 243, lane 7; purified GBS170d-His is shown in Figure 234, lanes 5-6.

Example 2103

A DNA sequence (GBSx2218) was identified in *S. galactiae* <SEQ ID 6505> which encodes the amino acid sequence <SEQ ID 6506>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45  bacterial cytoplasm --- Certainty=0.0502 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50 >GP:BAB06906 GB:AP001518 argininosuccinate synthase
   (citrulline-aspartate ligase) [Bacillus halodurans]
   Identities = 262/396 (66%), Positives = 321/396 (80%), Gaps = 1/396 (0%)

55 Query: 1  MGKELILAYSGGLDTSVALAWLK-KYDVIIVCMQDVGBGKLDLDFHDKALITGAIESTY 59
   M K+K++LAYSGGLDTSVAI WL K YDVIIV +DVGEGKL+*+ *KAL +GAIESY
   Sbjct: 1  MSKKKVLVAYSGGLDTSVAIKWLSDRGVIYIVAGLDVGBGKLELFPKSKALKVGAIESTY 60

60 Query: 60  LDVKEDEAEHFVLPALQNAHYBQKYPVLSALSRPLIAQKLVEMAHQTGATTINGCTGK 119
   +D K EFAE FYLPALQNAH+YBQKYPVLSALSRP+I++KIVFAA QTGA +RNGCTGK
60 Sbjct: 61  IDAKKFAEELVLPALQNAHYBQKYPVLSALSRPLISKLVEIAEQTGAQVAHQGCTGK 120

```


-2375-

Query: 120 GNDQVRFEVAIALDPELKVIAPVREWMHRESEITPAKANGVPIDADLNPFYSIDQNLN 179
 GNDQVRFEV+I AL+P L+V+APVREN W R+EEI +AK N +PIP DLNPFY+DQNLN
 Sbjet: 121 GNDQVRFEVSIQALNLEVLAPVREWMSRDEIEYAKNNIPIDLNPFYSVDQNLN 180

Query: 180 GRANEQGVLENPMQAPEERAGITKSPERAPDCABEYIDITFQNGKPIAINNQEMTLADLI 239
 GR+NEQ+LE+PW PE A+ +T + E+APD E +II F+ G P+ +N + +LI
 Sbjet: 181 GRNREGILEDPWATPPEGAYELIARIDAPDQPELVEIGPEKGIPTVLNKGSPVHELI 240

Query: 240 LSLNEIAGKHGIGRIDHYENRLVGIKREIYECPAWVLLAAHKRIEDLTLVREVSHPKP 299
 L LN+ LAGKHG+GRIDHYENRLVGIKRE+YECP AM L+ AHKE+EDLTL +EV+HPKP
 Sbjet: 241 LELNQLAGKHGVGRIDHYENRLVGIKREVTECPAMTLIKAHKELEDLTLTKEVAHPKP 300

Query: 300 ILFNKLSNLIYNALWPSATKAILIAYVKETQKVNGITTKVKLYGSAQVVAHRSNSLYD 359
 +E +++ LIY LWSPF A+ A++KETQ V G +VKL+KG A V R S SLY+
 Sbjet: 301 VVERKIAELIYEGLWPSFLQALSAFLKKTQSTVTGVVRVKLFKGAHVEGRKSESLYN 360

Query: 360 ENLATYTTAASFDQDAANGFIKLWGLPTQVIAQVNM 395
 E LATYT D FD +AANGFI LWGLPT+V + VNM
 Sbjet: 361 EKLATYTPDDEFDRNAANGFISLWGLPTKVYSVMNK 396

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2104

A DNA sequence (GBSx2219) was identified in *S.agalactiae* <SEQ ID 6507> which encodes the amino acid sequence <SEQ ID 6508>. This protein is predicted to be argininosuccinate lyase (argH). Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2131(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06905 GB:AP001518 argininosuccinate lyase [Bacillus halodurans]
 Identities = 284/454 (62%), Positives = 350/454 (76%)

Query: 6 KLWGRFPBSSLEKWEPEGASISFDQKLAFYDMKASMAHVIMLGKTDIISQBEAGLIKDG 65
 KLWGRFP + E WV+EPGASIFDQ+L D++ S+AHVIML K+ I++ EE IK G
 Sbjet: 3 KLWGRFPKTAQAWDEFGASIGFDQQLVEEDIESSLAHVIMLEKSGILANIEVQIKDG 62

Query: 66 LKILQDKYRAGQIFPSISNEDIHNMISLLTARIGEVAGKLTARSNDQVATWHLYLK 125
 L IL +K + G+L +S++NEDIH+NIE LL EIG V GLKLT RSRNDQVATWHLYL+
 Sbjet: 63 LHLLEKAKKGLNYSVANEDIHNLTKILLIDIGFVGGKLTGRSRNDQVATWHLYLR 122

Query: 126 DKLCBMKKLLHLRTTLWNLAEHHIYTVNPGYTHLQRAQPISPGHILMAYNNP+RUTER 185
 +E+++ + +++ IW A+ H+ T++PGYTHLQ AQPISF HIL+AY+ M RD R
 Sbjet: 123 KQTKELLQLVKNVQALVDAQKHVEITLIPGYTHLQRAQPISPAHILAYFWMLEKDYGR 182

Query: 186 LEFNKHTNLSPLGAALAGTTFPIRDHMTTLIDFEKPYSSNLDVARSURDIFIEPLSNA 245
 E +K+ N+SPLGA ALAGTTFPIR T LL F+ Y NSLDVSRDIFIEPLS+
 Sbjet: 183 YEDSLKRLNYSPLGAGALAGTTFPIRDREYTLILGFDGIYNSLDVARSURDIFIEPLSAS 242

Query: 246 SILMHLARPCBEILNWCSEYEQFITLSDTFSGSSIMPQKNPNMRLIRKTKGRVYGN 305
 S+LM HLAR CBE+I M S E+QF+ + D F+SGSSIMPQKNPNMRLIRKTKGRVY+G
 Sbjet: 243 SILMTHLSRLCKEILNWSQEPQFVEMDIAFATGSSIMPQKNPNMRLIRKTKGRVYGS 302

-2376-

Query: 306 LPSLLTVMKSLPLAYNKDLQEDKRGMPDSVETVSAIEIMANMLETVMNEHIMMTSTET 365
 LPSLLTV+K LPLAYNKD+QEDKRGMPD+V+TV ++ I A M++TM V E M +
 Sbjct: 303 LPSLLTVLKLGLPLAYNKDMQEDKRGMPDAVKTVKGSIAIFAGNIQTMVKESTMTKAVHQ 362

5 Query: 366 DFSNATELADYLASKGVFFRKAHEIVGKLVLECSKNGSYLQDIPLYYQBISELIENDIY 425
 DFSNATELADYLA+KG+PFR+AHK+VGKLV L C + G YL D+PL Y+ S+L + DIY
 Sbjct: 363 DFSNATELADYLATGMMPFREAEHVVGKLVLLCTIQRSIYLLDLPLSDYKAASDLFDEDIY 422

10 Query: 426 EILTAKTAVKRNSLGSTGTFQVKKILLARKEL 459
 ++L KT V RR S GGTGF +VK I A K L
 Sbjct: 423 DVLQPKTVVARRTSAGGTGFTVKKAIKAEEKIL 456

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 2105

A DNA sequence (GBSx2220) was identified in *S. agalactiae* <SEQ ID 6509> which encodes the amino acid sequence <SEQ ID 6510>. This protein is predicted to be class-II aldolase (fba). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2930 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9289> which encodes amino acid sequence <SEQ ID 9290> was also identified. Analysis of this sequence reveals:

30 GVH: Signal Score (-7.5): -2.92
 Possible site: 42
 >>> Seems to have no N-terminal signal seq.
 ALOM program count: 0 value: 0.37 threshold: 0.0
 PERIPHERAL Likelihood = 0.37 66
 35 modified ALOM score: -0.57
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2930 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP: BAB16889 GB: AB050113 class-II aldolase [Streptococcus bovis]
 Identities = 221/242 (91%), Positives = 234/242 (96%)

50 Query: 1 MAIVSAEKFVQARDNGYAVGGPNYNLEWTQAILRAAEAKKAPVLIQTSMGAQKMGY 60
 MAIVSAEKF++AAR+NGYAVGGPNYNLEWTQAILRAAEAKKAP+LIQTSMGAQKMGY
 Sbjct: 1 MAIVSAEKFIKAARENGYAVGGPNYNLEWTQAILRAAEAKKAPVLIQTSMGAQKMGY 60

Query: 61 KLCKQLIETLVESMGITVPVVAHLHDHGHYDALECEIEVGYSIMPDGSHLPVERNLEKAR 120
 KLCK LIE LVESMGITVPVVAHLHDHGH++DALECEIEVGYS+MPDGSHLPVERNLEK+
 55 Sbjct: 61 KLCKLTIEENLVESMGITVPVVAHLHDHGHFDALCEIEVGYSVMPDGSHLPVERNLEKAR 120

Query: 121 EVVAKAHAKGISVRAEVTGTGGEEDGIVGGELAPIEDAKAMVETGIDFLAAGIGNIHGP 180
 EVVAKAHAKG+SVRAEVTGTGGEEDGIVG GELAPIEDAKAMV TGI DFLAAGIGNIHGP
 Sbjct: 121 EVVAKAHAKGVSVRAEVTGTGGEEDGIVGGELAPIEDAKAMVATGIDFLAAGIGNIHGP 180

-2377-

Query: 181 YPANWEGILDHLKCLTAAVVGFPFIVLHGSGIPDDQIQEAIKLGAKVKNVNTBQQLAPC 240
 YPANW+GL LDHLKCLT AVVGFPFIVLHGSGIPDDQI+ AIKLGAKVKNVNTBQ+AF
 Sbjct: 181 YPANWQGLHLDHLKCLTAAVVGFPFIVLHGSGIPDDQIKAAIKLGAKVKNVNTBQQLAPC 240

Query: 241 QA 242
 +A
 Sbjct: 241 KA 242

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6511> which encodes the amino acid sequence <SEQ ID 6512>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2930(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 217/242 (89%), Positives = 228/242 (93%)

Query: 1 MATVSAEKFVQAAARDNGYAVGGFNTNNLEMTQAILRAAEAKAPVLITQSMGAARYMGY 60
 MATVSAEKFFQAAAR+NGYAVGGFNTNNLEMTQAILRAAEAK+APVLITQSMGAARYMGY
 Sbjct: 1 MATVSAEKFVQAAARENGYAVGGFNTNNLEMTQAILRAAEAKAPVLITQSMGAARYMGY 60

Query: 61 KLCQQLIETLVKSMGITVPVAIHLDHGHYDALECIYGYTSMFQSHLPVEENL 120
 K+C+ LI LVESMGITVPVAIHLDHGHY+DALECIYGYTSMFQSHLPVEENL K
 Sbjct: 61 KVCQQLITNLVSMGITVPVAIHLDHGHYDALECIYGYTSMFQSHLPVEENLAKTA 120

Query: 121 EVVAKAHAKGISVBAEVGTIGGEEDGIVGQELAPIEDAKAMVETGIDFLAAGIGNIHGP 180
 EVV APAG+SVBAEVGTIGGEEDGI+GQELAPIEDAKAMVETGIDFLAAGIGNIHGP
 Sbjct: 121 EVVKIAHAKGVSBAEVGTIGGEEDGIIIGQELAPIEDAKAMVETGIDFLAAGIGNIHGP 180

Query: 181 YPANWEGILDHLKCLTAAVVGFPFIVLHGSGIPDDQIQEAIKLGAKVKNVNTBQQLAPC 240
 YP NWEGIL DLHL+KLIT AVPGFPFIVLHGSGIPDDQI+EAI+LGAKVKNVNTB Q+AF
 Sbjct: 181 YPENWEGILDHLEKLTAAVVGFPFIVLHGSGIPDDQIKAIRLGAKVKNVNTBQQLAFS 240

Query: 241 QA 242
 A
 Sbjct: 241 NA 242

SEQ ID 9290 (GBS683) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 8 & 10; MW 55kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 11-13; MW 30kDa) and in Figure 184 (lane 11; MW 30kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2106

A DNA sequence (GBSx2221) was identified in *S.agalactiae* <SEQ ID 6513> which encodes the amino acid sequence <SEQ ID 6514>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2775(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2378-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAA88585 GB:M18954 unknown protein [Streptococcus mutans]
 Identities = 109/229 (47%), Positives = 156/229 (67%), Gaps = 1/229 (0%)

Query: 1 MFSGKRLKRRRTITAGYSQSELDKIHNRSSYPNWEKTKPNQNSLKQLAILLDVPEY 60
 MFS ++LK+RR IG SQ++ ADEL I+R SYPNWE KTKPNQ NL +LA LL V Y
 10 Sbjct: 1 MFSSQKLRKRRRKILGHSAQTADKLGISRSPSYNWEIGTKTKPNQNLKLAHLKVDASAY 60

Query: 61 FSESEKIVNTYLQSLQKQKVEKYABELLQTKVHEKIVPLFAVEVLSBIQLSAGPGBG 120
 F S++ IV Y +L+ N+ K KY++ LL+ Q ++ +LSAG G
 15 Sbjct: 61 FLSQHDIVEITYRLNESNKTILKYSQHLLQODKKRNLNKNKRYPIRYEKLASGTGYS 120

Query: 121 LYDEFETSETVSEDEYTGFDIAIWISGNSMEPVYKDGSEVALIRSTGFDHGDVAITANNNG 180
 + + +TV+ ++E D A+WI G+SMKP++ +GEVALI+ TGFD+DGA+YA++W+G
 15 Sbjct: 121 YFGDGNFDYTFVDEED-HDFASWIPGDSMEPTFLNGSEVALIKQTGFYDGAITYAIDWDG 179

Query: 181 SLYIKKLYRKEDGRFVMSINPDVAERFIPFEDSIRIVGKIVGHFMPVIG 229
 YTKK+YRKE G R+VS+N A++F P+++ RI+G IVG+F+P+ G
 20 Sbjct: 180 QTYIKKYREETGLRLVSLNKYADKAPYDENPRIIGLTVGNFPLEG 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6515> which encodes the amino acid sequence <SEQ ID 6516>. Analysis of this protein sequence reveals the following:

25 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4340 (Affirmative) < succ>
 30 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/209 (40%), Positives = 130/209 (62%), Gaps = 9/209 (4%)

35 Query: 25 LHNRSYPNWEKTKPNQNSLKQLAILLDVPEYFSESEKIVNTYLQSLQKQKVEK 84
 LH+N+ + NWE K PN+ +L L L +V YP+ Y+++ Y QL++ N+EKV
 Sbjct: 5 LFNVMKTIISNWEKQKIHNEKHLNALLHLFNVTSYDFPNYRLITPYNQLTISNKEKIVG 64

40 Query: 85 YABELLQQTQ-----KVHEKIVPLFAVEVLSBIQLSAGPGBGLYDEFETSETVSEDEYTG 138
 Y+E LL Q + +K L+A V LEAG G + + + V+ DE
 Sbjct: 65 YSERLLNHQIDKSKDLIDKPSQLYAYRYVES--LSAGTYSYFGDGNFDVVFY-DEGLE 121

45 Query: 139 FDIATWISGNSMEPVYKDGSEVALIRSTGFDHGDVAVYALNNGSLYIKKLYREEDGRFVMS 198
 +D A++W+ G+SMEP Y +GEV LI+ FD+DGA+YA+ N+G YIKK++R++G R+VS
 Sbjct: 122 YDFASWVPGDSNRPITYNGRVVLIQNPSFDYDGAITYAVENDGQTYIKKVPREDSEGLRLVS 181

Query: 199 INPDVAERFIPFEDSIRIVGKIVGHFMPV 227
 +N +++F P+ +E RI+GKI+ +F P+
 50 Sbjct: 182 LNKKYSDFAPYSEKPRIIGKILANRPL 210

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2107

55 A DNA sequence (GBSx2222) was identified in *S.agalactiae* <SEQ ID 6517> which encodes the amino acid sequence <SEQ ID 6518>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

-2379-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2387 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 2108

A DNA sequence (GBSx2223) was identified in *S.agalactiae* <SEQ ID 6519> which encodes the amino acid sequence <SEQ ID 6520>. This protein is predicted to be UmuC MucB homolog (uvrX). Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2195 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9925> which encodes amino acid sequence <SEQ ID 9926> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25

>GP: AAC98439 GB:L29324 UmuC MucB homolog [Streptococcus pneumoniae]
 Identities = 303/436 (69%), Positives = 360/436 (82%)

Query: 39 LHTSLCVMSRADNSAGLILASSPMFKKVFGRGNVGRAYDLFDVHTKRFNYYRAKISGLP 98

30

Sbjct: 5 LRLRLCVMSRADNSAGLILASSPMFKKVFGRGNVGRAYDLFDVHTKRFNYYRAKISGLP 64

Query: 99 TDAKFVSFIEWAKRTFIVPRMDLYIQNLEIQKVFQNYADPTDILPYSIDEGFIDILTS 158

T +V +IE WAK T IVP L I N-EIQK+PQ++A P DI PYSIDEGFIDILTS

Sbjct: 65 TTIDYVRYIEWAKRTFIVPREWILFIANMEIQKIFQDFAAPDDIYPYSIDEGFIDILTS 124

35

Query: 159 SLAYFVDEKSLSRKDKLDVVSAKIQHDIWKTGVYSTVGMNSANPLAKIALDNEAKTTA 218

SLAYFV DKS+SRKDKLD++SA IQ IW KTG+YSTVGMNSANPLAKIALDNEAK T

Sbjct: 125 SLAYFVDEKSLSRKDKLDIISAAIQKILWKTGVYSTVGMNSANPLAKIALDNEAKTTP 184

40

Query: 219 TVRANWSYEDVETKVMNIPKMTDFWIGISRTKRLNKLGIYSIKELANDPTILKKEKGV 278

TVRANWSYEDVE KVM IPKMTDFWIGIG+R EKRL+ LGL+SIKELA +P ++KKE G+

Sbjct: 185 TVRANWSYEDVENKVMNIPKMTDFWIGISRTKRLNKLGIYSIKELAQNPDLIAKKEG 244

45

Query: 279 IGVQHWFHANGIDESNVHEPYRPAKAVGIGNSQVLAHKDYTRQSDIELVLRMAHQVAILRL 338

+G++ WFHANGIDESNVH+PY+PK+ GIGNSQVLA KDY +Q DIE++LRMAHQVAILRL

Sbjct: 245 NGDELWFHANGIDRSNVHKKPKKSGIGNSQVLAHKDYTRQSDIELVLRMAHQVAILRL 304

50

Query: 339 RRRKKATVVAIVNGYSNPNKKSINQKINPNRNILVQDEVVSLFRSKYDGGAYRSIA 398

R KKATVVAI++GYS E K+SIN Q KI P N+T + + V+ L+ KY GA+R++A

Sbjct: 305 RSGKKATVVSILHSGYSNPNKKSINQKIEPTNQTALLTNYVLKLFHYTYSQAIRVA 364

Query: 399 VRVDGLVDENFAVISLFDDEFESEKKEKLETTIDSIDRNFGLAVQKASSILNKAISR 458

V Y GLVDE+P +ISLFD D E+ EKKE+L++ ID+IR FGF ++ K ++L + SR I+R

Sbjct: 365 VNYSGLVDESFGISLFDDEFESEKKEKLETTIDSIDRNFGLAVQKASSILNKAISR 424

55

Query: 459 SRLVGHSAGGLGGLK 474
 S+L+GGHSAGGL+GLK

-2380-

Sbjct: 425 SKLIQGHSAQGLDGLK 440

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2109

A DNA sequence (GBSx2224) was identified in *S.agalactiae* <SEQ ID 6521> which encodes the amino acid sequence <SEQ ID 6522>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4016(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2110

A DNA sequence (GBSx2225) was identified in *S.agalactiae* <SEQ ID 6523> which encodes the amino acid sequence <SEQ ID 6524>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2088(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAG13001 GB:AF227520 unknown [Streptococcus pneumoniae]
Identities = 68/122 (55%), Positives = 89/122 (72%), Gaps = 6/122 (4%)

35 Query: 1 MIDRSYLPFKVAREYQDRKMAQWGFPLSBHTAGLDSELNKVDYTSLSISDKLLLNQL 60
MIDRSYLPF+ AREYQD KM KWMGFPLSBHT+ L + NKV Y S+LS+ KLLLL+Q+
Sbjct: 1 MIDRSYLPFQSAREYQDTKMKQWGFPLSBHTSALTDANKVTYMSDLSLEKLLLSQV 60

40 Query: 61 YSNQLNGIIAVPGQ---YYBCKVDNLTFNHVSLEKTKTFVSIPIKDILSIDL--EVEYS 114
Y+ QLN I V + Y+Q + +LT + + +ET TG +++ +KDI+SI+L EV YE
Sbjct: 61 YAGQLNRIHVVKRWQVSYTGTLPSLTGDFILKTTTGHINIKAKDIVSIELVSEVLXE 120

Query: 115 SA 116
SA
45 Sbjct: 121 SA 122

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2381-

Example 2111

A DNA sequence (GBSx2226) was identified in *S.agalactiae* <SEQ ID 6525> which encodes the amino acid sequence <SEQ ID 6526>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 48
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4025 (Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9927> which encodes amino acid sequence <SEQ ID 9928> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2112

20 A DNA sequence (GBSx2227) was identified in *S.agalactiae* <SEQ ID 6527> which encodes the amino acid sequence <SEQ ID 6528>. This protein is predicted to be soluble transducer HtrXIII. Analysis of this protein sequence reveals the following:

```

      Possible site: 56
      >>> Seems to have no N-terminal signal sequence

25     ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5246 (Affixwative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2113

35 A DNA sequence (GBSx2228) was identified in *S.agalactiae* <SEQ ID 6529> which encodes the amino acid sequence <SEQ ID 6530>. Analysis of this protein sequence reveals the following:

```

      Possible site: 60
      >>> Seems to have no N-terminal signal sequence

40     ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5131 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2382-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2114

A DNA sequence (GBSx2229) was identified in *S. agalactiae* <SEQ ID 6531> which encodes the amino acid sequence <SEQ ID 6532>. This protein is predicted to be pXO2-78. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2105 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF13682 GB:AF180935 pXO2-78 [Bacillus anthracis]
Identities = 101/314 (32%), Positives = 147/314 (46%), Gaps = 46/314 (14%)

Query: 27 SGCIYEHDPDHSFRIFADTNTFKWFSRDIQGDVIDFVLVAGVSFKIALSYLETG--GFR 84
S + Y +HDS I N P W SR + G+I FVQ V SP A+ L G +E
Sbjct: 39 SERYRLTEHSDLIIDRKKQPYWNSRGVNGNIIKFVQVEDASFPGANQRLLDGEQDYE 98

Query: 85 EAKVIEETYPQFYLLREEP----PQARTYLKDIRGLSNQTINSFGRQGLLAQATYQAE 140
+A I +P+ Y E+ F +AR YL + R + Q +++ +GH+ Q Y
Sbjct: 99 KASEITFVSEPDYEHFQKEVSRPDAREVYLIEERKIDPQVVDALHNKGLIKQDKY-- 156

Query: 141 SVLVFKSFDENGLTQAASLQGLVQNERKYDRGLYKIMKSGSHQVGLSPDIGNPKRLIFC 200
+VL G + S QG+VK++ KY RG K I K S + G + G P+ L F
Sbjct: 157 NVLFWKDRDTQAVMGSGSEQGVKSD--KYRGAWKSQKSTANYGNVNLGSPRNLFKY 215

Query: 201 ESVIDMMYYQLHQKQLSDVRLISMGLKLSVIAVQTLLRLAEEQGLAFLDTVKPIRLS 260
BS ID++SY LH+ L D LISMEGLK VI +
Sbjct: 216 ESDIDLLSYATLHKHNLKDTHLISMGLKPQVI-----FN 250

Query: 261 HYLQAIQETTTFFQTNSNVITMAVNDDEAGREFYQKL-----SDKGFPIFQ-DLPLQ 312
Y++A + + +++ VND+AQ+ F ++L +D F+ + P
Sbjct: 251 YTMKACERIGDV---PDSLSLVDNDKAGKAFVERLIHFRYKEDGSIVAFKPEYPOAP 306

Query: 313 RLETKSDNDIVKR 326
E K DND KR
Sbjct: 307 SEEKNDNDNDCKR 320

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2115

A DNA sequence (GBSx2230) was identified in *S. agalactiae* <SEQ ID 6533> which encodes the amino acid sequence <SEQ ID 6534>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.7013 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2116

A DNA sequence (GBSx2231) was identified in *S.agalactiae* <SEQ ID 6535> which encodes the amino acid sequence <SEQ ID 6536>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1310 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2117

A DNA sequence (GBSx2232) was identified in *S.agalactiae* <SEQ ID 6537> which encodes the amino acid sequence <SEQ ID 6538>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.6726 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9373> which encodes amino acid sequence <SEQ ID 9374> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2118

A DNA sequence (GBSx2233) was identified in *S.agalactiae* <SEQ ID 6539> which encodes the amino acid sequence <SEQ ID 6540>. This protein is predicted to be phosphoglucomutase (manB). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2147 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9355> which encodes amino acid sequence <SEQ ID 9356> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB96418 GB:AJ243290 phosphoglucosyltransferase [Streptococcus thermophilus]
 Identities = 391/465 (84%), Positives = 424/465 (91%), Gaps = 1/465 (0%)

15 Query: 1 MAQHGIKSYVFEALRPTPELSFAVRHINAYAGIMVTASHNPAPFNGYKVGQDGGQLPPA 60
 + A HGIKSYVFE+LRPTPELSFAVRHL+ +AGIM+TASHNPAPFNGYKVG+DGGQ+PPA
 Sbjct: 107 LAARGIKSYVFEALRPTPELSFAVRHRLHTFAGIMITASHNPAPFNGYKVGEDGGQMPA 166

 Query: 61 DADALTDPIRAIENPFPAVELADLDESKSGSLIQVIGEDVDIYLRVIDVDNINQDLINFP 120
 DADALTD+IRAI+NPF V+LADL++SK+SGSLI++IGE+VD EYL+EVIDVDNINQDLIN +
 Sbjct: 167 DADALTDYIRAIIDNPTTVKADLDESKSGSLIEIIGENVDAEYLRKEVIDVDNINQDLINSEY 226

20 Query: 121 GDKMKIVYTFLEHGTGEMLTRKALAQGFBSVVVVSSQAKADPDFSTVSKSNPESQAQAFAL 180
 +GDKMKIVYT LHGTGEMLRKALAQGF++V VVE+QA DF TVKSRPE+Q AFAL
 Sbjct: 227 GRDMKIVYTSLEHGTGEMLRKALAQGFSDAVVQEAQAVPEADFLTVKSNPESQAQAFAL 286

25 Query: 181 AEELEGREVDADVLVATDPDADRLGVEIRQPDGSGYKNSGNQIGAIITAKYILEAHKTAGTL 240
 AEELEGRE VDADVLVATDPDADRLGVEIRQPDGSGY NLSGNQIGAIITAKYILEAHKTAGTL
 Sbjct: 287 AEELEGREVDADVLVATDPDADRLGVEIRQPDGSGYKNSGNQIGAIITAKYILEAHKTAGTL 346

30 Query: 241 PENAALAKSIVSTELVTKIABSYGNIMFNVLTGFKPIAEKIQEFSEKINHTYMPGFRESF 300
 P NAAL KSIIVSTELVTKIABSYGNIMFNVLTGFKPI EKI EPE +HN+TMFGFRESF
 Sbjct: 347 PANAALAKSIVSTELVTKIABSYGNIMFNVLTGFKPIGEKIEHEFTQGNHTYMPGFRESF 406

35 Query: 301 GYLIKPFVRDKDAIQAVLLVASTAAYRSRGLTADGIDEIYKGYFYPAKTIISVTLSGV 360
 GYLIKPFVRDKDAIQAVLL+VASTAAYRSRGL+TADGI+EIYK+YGYF+EKTIISVTLSGV
 Sbjct: 407 GYLIKPFVRDKDAIQAVLLVASTAAYRSRGLTADGIEIYKGYFYPAKTIISVTLSGV 466

40 Query: 361 DGAAEIKKIMDKFRENGPKQFNNDIVLLEDFQKQTATKNDGTISMLTTPSPNVLYKTLA 420
 DGAAEIKKIMDKFR N FQFNNDIDI EDF +QTAT DG + LTTPSPNVLYK LA
 Sbjct: 467 DGAAEIKKIMDKFRENGPKQFNNDIAKTEDFLEKQTATATDAG-VKLTTPSPNVLYKTLA 525

45 Query: 421 DDSWIAVRPSGTEPKIKFYIATVGNDLADAEKIANIEKRIITTFV 465
 DDSW AVRPSGTEPKIKFYIATVG ADA KIANIE SI FV
 Sbjct: 526 DDSWIAVRPSGTEPKIKFYIATVGSTEDAEKIANIEASINAFV 570

45 There is also homology to SEQ ID 6156:

 Query: 1 MAQHGIKSYVFEALRPTPELSFAVRHINAYAGIMVTASHNPAPFNGYKVGQDGGQLPPA 60
 +AQHGIKSYVFEALRPTPELSFAVRHINAYAGIMVTASHNPAPFNGYKVGQDGGQLPPA
 Sbjct: 107 LAQHGIKSYVFEALRPTPELSFAVRHINAYAGIMVTASHNPAPFNGYKVGQDGGQLPPA 166

50 Query: 61 DADALTDPIRAIENPFPAVELADLDESKSGSLIQVIGEDVDIYLRVIDVDNINQDLINFP 120
 DADALTDPIRAIENPFPAVELADLDE+KSGSLIQVIGEDVD+SYLRVIDVDNINQDLINFP
 Sbjct: 167 DADALTDPIRAIENPFPAVELADLDEKNSGSLIQVIGEDVDIMRYLRVIDVDNINQDLINFP 226

55 Query: 121 GDKMKIVYTFLEHGTGEMLTRKALAQGFBSVVVVSSQAKADPDFSTVSKSNPESQAQAFAL 180
 GDKMKIVYTFLEHGTGEMLTRKALAQGFBSVVVVSSQAKADPDFSTVSKSNPESQAQAFAL
 Sbjct: 227 GDKMKIVYTFLEHGTGEMLTRKALAQGFBSVVVVSSQAKADPDFSTVSKSNPESQAQAFAL 286

60 Query: 181 AEELEGREVDADVLVATDPDADRLGVEIRQPDGSGYKNSGNQIGAIITAKYILEAHKTAGTL 240
 AEELEGREVDADVLVATDPDADRLGVEIRQPDGSGYKNSGNQIGAIITAKYILEAHKTAGTL
 Sbjct: 287 AEELEGREVDADVLVATDPDADRLGVEIRQPDGSGYKNSGNQIGAIITAKYILEAHKTAGTL 346

 Query: 241 PENAALAKSIVSTELVTKIABSYGNIMFNVLTGFKPIAEKIQEFSEKINHTYMPGFRESF 300
 PENAALAKSIVSTELVTKIABSYGNIMFNVLTGFKPIAEKIQEFSEKINHTYMPGFRESF
 Sbjct: 347 PENAALAKSIVSTELVTKIABSYGNIMFNVLTGFKPIAEKIQEFSEKINHTYMPGFRESF 406

-2385-

Query: 301 GYLKPFVRDKDAIQAVLLVARIAAYYRSGLTADGIDEIKYKGYFARKTISVTLSGV 360
 GYLKPFVRDKDAIQAVLLVARIAAYYRSGLTADGIDEIKYKGYFARKTISVTLSGV
 5 Sbjct: 407 GYLKPFVRDKDAIQAVLLVARIAAYYRSGLTADGIDEIKYKGYFARKTISVTLSGV 466

Query: 361 DGAAEIKKIMDKFRNGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA 420
 DGAAEIKKIMDKFRNGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA
 10 Sbjct: 467 DGAAEIKKIMDKFRNGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA 526

Query: 421 DDSNIIVRPSGTETPKIKFYLATVGNLADAEKIANIEKEITTFV 465
 DDSNIIVRPSGTETPKIKFYLAT+G+ L A+ KIANIE EI TFV
 10 Sbjct: 527 DDSNIIVRPSGTETPKIKFYLATIGTILDAQKIANIETETTFV 571

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 2119

A DNA sequence (GBSx2235) was identified in *S.galactiae* <SEQ ID 6541> which encodes the amino acid sequence <SEQ ID 6542>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1564 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9905> which encodes amino acid sequence <SEQ ID 9906> was also identified. There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 2120

A DNA sequence (GBSx2236) was identified in *S.galactiae* <SEQ ID 6543> which encodes the amino acid sequence <SEQ ID 6544>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.92	Transmembrane	162 - 178 (135 - 184)
INTEGRAL	Likelihood = -7.11	Transmembrane	58 - 74 (56 - 78)
INTEGRAL	Likelihood = -6.42	Transmembrane	136 - 152 (135 - 161)
INTEGRAL	Likelihood = -5.20	Transmembrane	23 - 39 (21 - 49)
INTEGRAL	Likelihood = -1.75	Transmembrane	485 - 501 (485 - 501)

----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35376 GB:AB001710 ABC transporter, ATP-binding protein
 [Thermotoga maritima]
 Identities = 216/552 (39%), Positives = 336/552 (60%), Gaps = 3/552 (0%)

50 Query: 26 MALLSTGVQVCLTVYLPVLIGQAVDVLSPSMILLPIWMMIAVILANTYIQWINDPLL 85
 M + V L V P LIG+ +DVV P LL M + + +++W+ +

-2386-

Sbjct: 41 M V F V F V T V S S I L G V L S P L I G K T I D V F V P R R F D L L P R Y M L I G T I Y A L T S L F W L Q G K I 100

Query: 86 Y N R L I P H Y V A S L R K A V M K A N L L P I A Y L O K R G I G D L I S R V T I D T E Q L S N G L M V F N Q F F V 145
L V L R K + E K I + P + + D + G D + I S R V D + + N L Q F F

5 Sbjct: 101 M L T L S Q D V F R L R K E L Q R V F V G F F D R T P E G D I I S R V I N D V D N I N N V L G N S I I Q F F S 160

Query: 146 G L T I I V T F S M A K I D L M L F L V L F I T P L S L F L A R F I A K S Y - H Y Q N C T A S R G R Q T O F I 204
G + + T + + M + + + + + + L + P L + + + + + + + Y + N Q G + I

10 Sbjct: 161 G I V T L A G A V M F R V M V I L S L A V T L S I V P L A T V L I T Q I V S S Q T R K Y P Y E N Q R V L - G Q L N G I I 219

Query: 205 R E M V S Q E S L I Q A F S A Q E R S S D H P R T I N Q E Y A N F S Q S A I P Y S S T V M P S T R F I N S I Y G F L A 264
E E + S + + I + F + E + + F + N + A + S + P + N + L + + +

15 Sbjct: 220 K E D I S Q L T V I K L P T R E E K E M K F D R V N E S L R K V G T K A Q I P S G V L P P I M M V I N L G S P A L T S 279

Query: 265 G I G A L R I M S G A F S V Q Q L I T T L N Y V N Q Y T K P F N D I S S V L S E M Q S A L A C A E R L Y S I L E S S P 324
G G + + V G + T P + Y Q + T + P N + + S + + Q A L A A E R + + I L +

20 Sbjct: 280 G F G W L A L K D I T T V G T A T T I G Y S R Q F T R P N E L S N Q F M I Q M A L A S A E R I F E L L O S E E 339

Query: 325 N I T G T E K L D S S T V K Q I D F K N V F G Y N K S L L L G I N L H I P A G A K V A I V G P T G A G K S T L I 384
+ + V + G + I + F K N V Y + K K + L I H I G K V A + V G P T G + G K + T + +

25 Sbjct: 340 K - D D P A V E L R E V G E I F F K N W F S Y K K K P L K D I T P H T K P G Q K V A L V G P T G S G K T I T V 398

Query: 385 N L I M R F Y E V D G G N I L L D C K P I T D Y E P S Q L R Q S I G M V L Q S T W L K S A T I H D N I A Y N P K A S R 444
N L + M R F Y + V D G I L + D I + S L R I G + V L Q + T L S T + N + Y N P A +

30 Sbjct: 399 N L I M R F Y D V D G Q L V D G I D R K I R S S L R S S I G I V L Q T I L F S T T V K E N L K Y N G P A T D 458

Query: 445 E E V I S A A K A N A D F F I K Q L P N G Y D T Y L E D A G S L S Q S Q Q L L T I A R I F L K L P R I L L I D E A 504
E E + E A A K + + D F I K L P G Y + T L D G + L S Q Q Q Q L L I R F L P + I L L I D E A

35 Sbjct: 459 E E I K E A A K I T H S D F I K H L P S G Y E T V L T N G E D L S Q Q Q Q L A I T R A F L A N F L K I L L I D E A 518

Query: 505 T S S I D R T E V L Q E A P Q M L M K G R T S F I I A H R L S T I Q T A D I I L V M V S G E I V E M G N H S E M A 564
T S + D T + T E + Q A L M + G + T S I I A H R L + T I + A D + I + V + G E I V E + G H E L +

40 Sbjct: 519 T S V I D T K T E S I Q A A W K L M S G K T S I I L A H R L A T I N A D L I V L R D G E I V E M G K H S E L I Q 578

Query: 565 Q N G I Y Y Q M C N A Q 576
+ + G Y Y + + + Q

35 Sbjct: 579 K R G F Y I E L F T S Q 590

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6545> which encodes the amino acid sequence <SEQ ID 6546>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have an uncleavable N-term signal seq

45	INTEGRAL	Likelihood = -8.07	Transmembrane	162 - 178 (159 - 182)
	INTEGRAL	Likelihood = -7.17	Transmembrane	143 - 159 (137 - 161)
	INTEGRAL	Likelihood = -5.84	Transmembrane	23 - 39 (19 - 45)
	INTEGRAL	Likelihood = -5.68	Transmembrane	68 - 84 (60 - 86)
	INTEGRAL	Likelihood = -2.55	Transmembrane	261 - 277 (256 - 278)

----- Final Results -----

50	bacterial membrane	--- Certainty=0.4227 (Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000 (Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 >GP:AAD35376 GB:AE001710 ABC transporter, ATP-binding protein
[Thermotoga maritima]
Identities = 206/572 (36%), Positives = 342/572 (59%), Gaps = 5/572 (0%)

60 Query: 2 I K T D H I L L K R V L Q Q L L K K V L P V C I L V I A S F V Q V G -- L S V L P V L I G R A V M S L G V N S N Q T 59
+ K L + R + L L + P + + + + F V V L V P L I G K + D + +

Sbjct: 18 L K N P T A I L R R L L G Y L -- R P H T F T L I M V F V V T V S S I L G V L S P L I G K T I D V F V P R R F D L 75

Query: 60 L K W L G Q M I V I V N T L I Q W M P L V S R L L Y Q S Q Q L K L L E K I H R L P F A Y L D R Q T I G D 119
L + + + I + + L + W + + L + L + + L E K + R + P + D R G D

65 Sbjct: 76 L P R Y M L I G T I Y A L T S L F W L Q G K I M L T S Q D V F R L R K E L Q R V F V G F F D R T P E G D 135

-2387-

- Query: 120 LVSRVITVTEQLINGLQMVFNQFILGLLTILCTIIAMQIDMLMLLVLTVPSSFLAR 179
 ++SRVI D + + N L QF G++T+ +I M +++ ++ L + P ++ + +
 Sb|ct: 136 IISRVINDVDNINHVLSNLSIQFFSGIVTLAGAVIMFRVNVLSLVTLVSVLPLTLTQ 195
- Query: 180 FIAQKSPHYQAQCTKSRGNLAQFTEKILRQEGVLQVFNABQSCIDYHVNKTYCEABQK 239
 ++ ++ Y + G L EE + ++LF +E+ + + N++ + +
 Sb|ct: 196 IVSSQTKYFYENQVLSQLNGIIRDISGLTVIKLITREEKEMKKPDRVNESLRKVGTK 255
- Query: 240 AIFYASTVNPATRFINSVIYALLAGLAGVIRIAGLFSVQGLTTFINVVVQYTKPFNDISS 299
 A ++ + P +N++ +AL++G G + + +VG + TF+ Q+T+P N++S+
 Sb|ct: 256 AQIPSGVLPPLMMNVINLGFALISGPGMLALKDITVTGTIATPIGYSRQPTFPLNELSN 315
- Query: 300 VLAEIQSSLACARLYDLDDIRIKQBEHFILFKASAVKGQIDFEVSPSYQKDRPLKDI 359
 IQ +LA A+R++++LD+E +E++ + V+G+I+P+ V FSY K +P+LKDI
 Sb|ct: 316 QPNMIQMALASABRIFLIDLE-EEKDDPDVLELRVGRBIFKNWVFSYDKKKFVLKDI 374
- Query: 360 NFSVPAGS+KVAIVGPTGAGKSTLINLMRFYELDAGSIKLDKVPICKYAKBELRSTGIV 419
 F + G KVA+VSPGTG+GK+T+NLNMRFY++D G I +D + I+ + LRS GIV
 Sb|ct: 375 TPIHKGGKQVAVGPTGSGKTTIVNLMRFYDVRGQILVDGIDIRIKRSSLRSTGIV 434
- Query: 420 LQETWLKDATVHLLIYAGSEASRDEVVAAAKAAHAFPMQLPKTYDTYLSASDALSQ 479
 LQ+T L TV E + YG+ A+ +E+ AAK H+ FI LP+ Y+T L+ + + LQ
 Sb|ct: 435 LQDTILFSTTVKENLQNGGATDEEIKAAKLTASDHFIKHLPEGYSTVLTDMGDELQ 494
- Query: 480 GQLQLLAIRMLPLKKKLYVLDEATSSSIDIRTEAVIQEALKELMGRGTSPIAHLRSTIQ 539
 GQ QLAI R FL PK+L+LDEATSS++D +E IQ A+ +LM G+TS IDAHLR+TI+
 Sb|ct: 495 GQQLLAITRAFLANPKLILDEATSSVDITKESIQAMWKLMEGKTSIIDAHLRNTIK 554
- Query: 540 SADDLIVMDQSRIVEMGTHASLSNKGCTVRL 571
 +ADLI+V+ G +VE G H L+ K G Y L
 Sb|ct: 555 NADLILVLRDGEIVMGKDELQKRGFYEL 586

An alignment of the GAS and GBS proteins is shown below.

- Identities = 340/566 (60%), Positives = 433/566 (76%)
- Query: 11 KKLVDQLSKKSLVGMALLGTVVQVCITVYLPVLIGQAVDVLSPHSMILLPIMKMGIA 70
 K+++QDLL K V + ++ + VQV L+YLPVLIG+AVD+ LS +S L ++ ++
 Sb|ct: 10 KRLVQDLLKKPLPVCLIVIASFVQVGLSVYLPVLIGKAVDMSLVNSWQLTKWLLGQMLV 69
- Query: 71 VILANTIQINPLLYNKLPHYVASLRKAMKINLFIAYLDRKGIGDLISRVITDTE 130
 +I+ NT+IQW+ PL+Y+RL++ Y L+ ++EK++ LP AYLD++ IGLD+SRV TDE
 Sb|ct: 70 IIVVNTLIQWVPLVYSRLLYQNSQQLKDKLEKTHRLFPAYLDRQTIQELVSRVITDTE 129
- Query: 131 QLSNGLLMVFNQFFVGLLTIIVTIFSMKIDLLMLSLVLFITPLSLFIARFIKSYHYLY 190
 QL NGL NVFNQF +GLGTT+ TI +MA-ID LML LVL LTP SLFLARPIA+KS+H+
 Sb|ct: 130 QLINGLQMVFNQFILGLLTILCTIIAMQIDMLMLLVLTVPSSFLARPIAQKSPHYA 189
- Query: 191 CNQTASRGRQTQFIEVMVSQESLIQAFSAQERSSDHPRTINQYANFSAQIAFYSTVNP 250
 Q QT SRG QF EB++ QE L+Q P+AQE+S + +N+ Y SQ AIFY+STVNP
 Sb|ct: 190 CAQTKSRGNLAQFTEKILRQEGVLQVFNABQSCIDYHVNKTYCEASQKAIFYASTVNP 249
- Query: 251 STRFINSLIYGLAGLAGIRIMSGAFSVQGLITFLNYYVNYTKPFNDISSVLSMOSALA 310
 +TRFINS+IY LAG+GA+RIM+G FSVQGLITFLN Y QYTKPFNDISSVL+R+QS+LA
 Sb|ct: 250 AIRFINSVIYALLAGLAGVIRIAGLFSVQGLITFLNVVVQYTKPFNDISSVLARIQSSLA 309
- Query: 311 CARLYSLLESSPNITGTETKIDSGSTVKGQIDPFKNVVFGYNSKILLNGINLHPAGAKV 370
 CR+RLY +L+ +S VKGQIDF+ V F Y K + LL IN +PAG+KV
 Sb|ct: 310 CAQRLYLDLIDIEIKQBEHFILFKASAVKGQIDFEVSPSYQKDRPLKDLNFSVPAGSKV 369
- Query: 371 AIVGPTGAGKSTLINLMRFYEVDDGNILLDCKPITDYEPSQRLQEGIMVLQETWLKAT 430
 AIVGPTGAGKSTLINLMRFYEVDD G+I LD PI Y +LR G+VLQETWLK AT
 Sb|ct: 370 AIVGPTGAGKSTLINLMRFYELDAGSIKLDKVPICKYAKBELRSTGIVLQETWLKAT 429
- Query: 431 IHDNIAYANPKASREEVIEAANAANAFPIQLPNGYDTYLRDAGDSLQSQGQQLTIAR 490
 +H+ IAY + +ASR+EV+ AAKRA+A FFI QLP YDTYL +D+LSQQL QLL IAR
 Sb|ct: 430 VHELLIYAGSEASRDEVVAAAKAAHAFPMQLPKTYDTYLSASDALSQQLQLLAIR 489

-2388-

Query: 491 IFLLKPRILLDEATSSIDRTTEVLVQRAFOQLMKGRTSFIIAHLSTIQADADILVMDV 550
 +FLK P++L+LDEATSSID RTE ++QRA + LM+QRTSFIHRLSTIQ+AD+ILVM
 Sbjct: 490 MFLKKPKVLVLDATSSIDRTTEAVIQALKELMKGRTSFIIAHLSTIQASADLILVMDQ 549

Query: 551 GEIVRVONHSELMAQKGIYYQMNAQ 575
 G +VE G H+ LM++ G Y ++Q + +
 Sbjct: 550 GRUVENGTASIMSKNGCTVRLQKIE 575

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2121

A DNA sequence (GBSx2237) was identified in *S.agalactiae* <SEQ ID 6547> which encodes the amino acid sequence <SEQ ID 6548>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 26
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 20 bacterial cytoplasm --- Certainty=0.1099 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2122

A DNA sequence (GBSx2238) was identified in *S.agalactiae* <SEQ ID 6549> which encodes the amino acid sequence <SEQ ID 6550>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

- 30 Possible site: 37
 >>> Seems to have no N-terminal signal sequence
- | | | | |
|-------------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -13.69 | Transmembrane | 157 - 173 (130 - 182) |
| INTEGRAL | Likelihood = -10.88 | Transmembrane | 56 - 72 (49 - 77) |
| 35 INTEGRAL | Likelihood = -7.75 | Transmembrane | 239 - 255 (235 - 258) |
| INTEGRAL | Likelihood = -6.42 | Transmembrane | 133 - 149 (130 - 156) |
| INTEGRAL | Likelihood = -4.78 | Transmembrane | 271 - 287 (270 - 289) |
| INTEGRAL | Likelihood = -1.91 | Transmembrane | 20 - 36 (20 - 37) |
- 40 ----- Final Results -----
- bacterial membrane --- Certainty=0.6477 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35375 GB:AB001710 ABC transporter, ATP-binding protein
 [Thermotoga maritima]
 Identities = 196/570 (34%), Positives = 327/570 (56%), Gaps = 5/570 (0%)

- 50 Query: 1 MKRLTYFFKGYIKETIFGFLFKLLEASPELLVPIVIARMDITETIPGRDRSGLLIQIGLIF 60
 MK L Y K Y + PLF ++E +L P ++R+++DE I RGD S L+L+ G++
 Sbjct: 1 MKTLARYLKPYNIFAVLAPLPMVVEVICDLSPPTLLARIVDGGIARDPS-LVNLGTGLM 59
- Query: 61 FLAA-VGVVVVAITQYYSSKAAVGYTRQLTDLQYQVMSLKKKDRDELGTASLITRLTAD 119

-2389-

```

      + A + G V I ++S A+ + L DL++KV+S ++ T+SLITRLT D
Sbjct: 60 LIVALIGAVGIGICTVFASYASQNFQADLRDLFRKVLSPSISVNRKPHSTSLITRLTND 119

5 Query: 120 TPQIQTGLNQFLRLFLRAPIIVFGALINAFSISPSTIKWFLVNVVTLFIVFVMSRLNAP 179
      Q+Q + LR+ +RAP++ G I+RA SI+ L+ + ++ ++ ++ NP
Sbjct: 120 VTQLQLVWMLRLIVRAFLFVGGIVMAVSINVKLSVLIPLIPFVILLFWLTKGNP 179

10 Query: 180 IYLKIRTSTDYLVKLTQQLQGVVIRAFWQVDRSEAFNDINHYTNLQLKAGRLSSLV 239
      ++ KI+ STD + ++ R+ L GVRV+RAF + + E+E F N + + A L
Sbjct: 180 LFFKIQBSTDEWNRVRENLLGVVRVRAFRREYENRFRKANESLRSSIRISAFSLIVFA 239

15 Query: 240 TPLTFLVNVITIVVIMKGNLNIANHLLSQMLNIALINYLQILVELLQMTLMTVLSNQS 299
      FL +VN+ ++ ++W G + + N+ + G +A+ NYL+QI+ L+ + ++ ++
Sbjct: 240 LFLFIFIVNKMIAVLWFGVLRVNRWMEIGSIYATNYLMQIMPSLMMIGLIMFVRA 299

20 Query: 300 YISAKRIIAVF-EPFS-BIIDDKLEPKYSNKALEVQRMASFYPMSSSEKALSDITFSMVG 357
      SAKR++ V B++ E D+ L ++ + + F Y +++ LS + FS+ G
Sbjct: 300 SASAKRVLEVLANEKAIREADNALALFNVBGSFENVEFRYFENTDFVLGSVNFYSKPG 359

25 Query: 358 ETLGIGGTGSCKSTLIMLLHIYKVBQSDIDYHQKSPDTISWRTIARVFPQNAQLF 417
      ++G TSSCKSTL+HL+ + + G +++ + + R + VPQ LF
Sbjct: 360 SILVAVLGSTGCKSTLIMLLIPRLIDPGRGEVDELDVVRVTKLDLRGHISAVPQETVLF 419

30 Query: 418 KGTIRNHSLSGLGKVSSEKLTALTAQSDPVYKEDQQLDAPVSPGRNFSQGGQRRLT 477
      GTI+ HL G +++++ A+IQ DP+ D+ VE GRNFSQGG+QRL+
Sbjct: 420 SGTIKENKWRGREDATDDEIVEAAKIQIHDFTIISLPEYDSRVERGRNFSQGGQRRLS 479

35 Query: 478 IARALVQDKIPFLILDATSLDYLTARLFAIKTHFNQTMILLIVSORINSIQNADRII 537
      IARALV+ K LILD TS++D +TE R+ ++ I++Q+I + AD+IL
Sbjct: 480 IARALV- KPKVLILDCTSSVDPITEKRIIDGLKRYTGCTTFTITQKIPTALLADKIL 538

Query: 538 LLDKKGQKQGFNNHQSILAHNKVYKSYHSQ 567
      +L +GK GF H+ LL H K Y+ IY GQ
Sbjct: 539 VLEHGKVGAGFTHKELLECKPYREIYESQ 568

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6551> which encodes the amino acid sequence <SEQ ID 6552>. Analysis of this protein sequence reveals the following:

```

      Possible site: 37
>>> Seems to have no N-terminal signal sequence
40 INTEGRAL Likelihood = -12.47 Transmembrane 157 - 173 ( 149 - 185)
      INTEGRAL Likelihood = -7.75 Transmembrane 55 - 71 ( 51 - 74)
      INTEGRAL Likelihood = -4.25 Transmembrane 239 - 255 ( 237 - 260)
      INTEGRAL Likelihood = -3.77 Transmembrane 20 - 36 ( 19 - 37)
      INTEGRAL Likelihood = -3.50 Transmembrane 271 - 287 ( 270 - 288)
45 INTEGRAL Likelihood = -2.55 Transmembrane 133 - 149 ( 130 - 151)

----- Final Results -----
      bacterial membrane --- Certainty=0.5989 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

1GB:AL137187 putative ABC transporter [Streptomyces ... 296 6e-79
55 >GP:CA669751 GB:AL137187 putative ABC transporter [Streptomyces
      coelicolor A3(2)]
      Identities = 185/569 (32%), Positives = 306/569 (53%), Gaps = 8/569 (1%)
Query: 1 MKRLRPYVKGYLKESIGLPLKLEALFELLVPLLIANNIDISISQNSGGLRVVLTIF 60
      ++ LR Y++ Y K L + L+ L + P L A+ID + + +8 IL +
Sbjct: 3 IRLRLTYLRPKKPIALLVALQFLTCASILPLFLANHIIDSGVVKESGYTILSYGALMI 62

60 Query: 61 GLATIGLLSVTAQYPSKKAAGVFTRQMTDILPKKIMFLSKEDQDHLQYASLIRITSDS 120
      G++ ++ ++ A ++ ++ A R+ +F ++ S + H G SL++R T+D
65 Sbjct: 63 GISLAQVVCNIGAVFYTGAKTAALGRDVRGAVDRVQSPSAREVGHGPAISLITTTNDV 122

```

-2390-

Query: 121 FQIQTGINQFLRLFLRAPIIVCGAMVAYWISPSLITMFMVMVIVLLTLVFMVSHLLGPL 180
 Q+Q L + AFI+ G +VMA + L+ + +V VL V++ L PL
 Sbjet: 123 QQQQMLAIAMFTTAMVASAPIMCVGGIVNALGLDVPLSGVLIGVVPVLAICVTYLIVRLRPI 182

Query: 181 YLLIRRETDLHLVRLTSQQQLQIRVIKAFNQTKRLQAFKQNMLLSRHQYQAANTLVANLV 240
 ++ D + R+ +Q+ G RVI+AF + + R Q P+ + N L+ L +
 Sbjet: 183 FRIMQVRLDTNVRVRLRQGTIGNRVIRAFRVEDRYEQRPRIKAMTELTEVALGTGMLLAIANF 242

Query: 241 PWTFLVNNITLLILIMQSSQVAHRSLSQGMVALINYLQILAEILLKMTLMATINQSV 300
 P+ VVNL+ + +W G + + + G L A + YL+QI+ ++ T + + ++
 Sbjet: 243 PVMVTVNLSIAVWFGAHRIDGGMQIGDLTFLAYLMIQIVMSVMATFMPFMVPRAE 302

Query: 301 TAAKRINQVFLADRAPLPLKDGPISTH-LITIRHLTFTYPGAAPSLDYQLSDAQGE 359
 A+RI +V P+ + H L IR F YPGA EP L I L A GE
 Sbjet: 303 VCAERIQEVLETESVVPVAPVTELRHGLRIRGFRYPGAEPSVLRHLDLVARPGE 362

Query: 360 WIGIIGGTGAGKTTLLIDLCQYYSQSGEISLNN---QGEVPKLTETWRNVIALVTPQKQ 416
 +IG TG+GK+TL+ L+ + + GE+ +N + PKTL + V++LVPKQ
 Sbjet: 363 TTAIVGSTGSGKSTLLGLVPLRFDATDGEVLVNGVDRVTDPKILAK---VVSIVPQRY 419

Query: 417 LFKGTIRSNLLGQSMPISEELWRALELAQAKEPVAALPEQLAPVEAPGRHPSGGQR 476
 LF GT+ +NL G + +DEEM AL +AQAKEF+ L L+AP+ G + SSGQR
 Sbjet: 420 LFAGTVATNIRYG-NPDATDEELWHALAQAQKEPVSELGGLDAPLAQGGTVNVSQGR 478

Query: 477 RLAIARALLKPKPILIDRASSALDNETGRQLPKALKEELSDVLVLTQTSKMLQFADK 536
 RLAIAR L++ I + DD+ SALD T L L +E ++ V++V Q + ++ AD+
 Sbjet: 479 RLAIARTLVQRPETLYLFDOSFALDYATDAALRAELAQETAEATTVIVAQRVATTEDADR 538

Query: 537 ILVLEQGHQOLDFASHDQLKVSNAIQEML 565
 L+VL++G + H +L N Y+E++
 Sbjet: 539 IIVLDGRVVGVRHMLNADNETYREIV 567

An alignment of the GAS and GBS proteins is shown below.

Identities = 313/568 (55%), Positives = 428/568 (75%), Gaps = 9/568 (1%)

Query: 1 NKRLTYYPKGYIKETI FGPLFKLEASPELLVPIVIAIMDETIPGRDREGILLQIGLIF 60
 NKRL Y KGY+KE+I GPLFKLLRA FELLVP++IA MID +I + + G+L + +F
 Sbjet: 1 NKRLRPYKGYLKESILGPLFKLEALFELLVPELLIANNIDTISISQHSNGILLVVLTLF 60

Query: 61 FLAAVGVVVAITAQYSSKAAVGYTRQLTEDLYQKVMGLKGRDDELGTASLIRLITADT 120
 LA +G++++TAQY+SSKAAV+GTRQ+T+DL++K+M L K D+D IG ASL++RLT+D+
 Sbjet: 61 GLATIGLLSVTAQYFSSKAAVGTQMTDOLFKKIMFLSKEDQHLGYASLLGRITSDS 120

Query: 121 FQIQTGANQFLRLFLRAPIIVGAIINAFSPISPSLITWFLVMVITLFIIVFMVSHLLNPI 180
 FQIQTG+NQFLRLFLRAPIIV GA++MA+ ISPSLT+WE++MV+ L +VFMV LL P +
 Sbjet: 121 FQIQTGINQFLRLFLRAPIIVCGAMVAYWISPSLITMFMVMVIVLLTLVFMVSHLLGPL 180

Query: 181 YLKIRTSYDYLVLKTRQQLQGVVIRAFNQVDRESAPNDINHYTHNLQKAGRLSSVT 240
 YL IR TD+LV+LT QQLQG+RVI+APNQ +E +AP N + Q +A L++++
 Sbjet: 181 YLLIRRETDLHLVRLTSQQQLQIRVIKAFNQTKRLQAFKQNMLLSRHQYQAANTLVANLV 240

Query: 241 PLTFLLVNNITLVVITWRGNINIANHLLSQGMVALINYLQILVELLKMVLVTSINQSV 300
 P+TFLLVN+TL+++IW+G+ ++ LSQGMVALINYLQIL ELLKMTM++ +NQS
 Sbjet: 241 PWTFLVNNITLLILIMQSSQVAHRSLSQGMVALINYLQILAEILLKMTLMATINQSV 300

Query: 301 ISAKRIIVAFV---ERPSEIIDDKLEPKYSNKALEVQEMAPSPYNSSEKALSITPSKNV 356
 +AKRI VF EP + + D S L ++ + P+YP ++E +L DI S +
 Sbjet: 301 TAAKRINQVFLADRAPLPLKDGPISTH-LITIRHLTFTYPGAAPSLDYQLSDAQGE 357

Query: 357 GETLGIIGGTGSGKSTLINLLHRIYKQSGDIDYHQGSKPDTISNWRITVLRVVPQKQL 416
 GE +GIIGGTG+GK+TL+L+ Y G+I + QG+ P T++ WR ++ +VQP Q+L
 Sbjet: 358 GEWIGIIGGTGAGKTTLLIDLCQYYSQSGEISLNNQGVKTLTETWRNVIALVTPKQQL 417

Query: 417 FKGITIRSNLSIGL-KVSEKILWALRIQAQSDFVKKKDGLQAVPFSKGNPSGGQRQR 475
 FKGITIRSNL IG +S+E+LW ALE+AQ+ +FV QL+APVE+PGR+PSSGQRQR
 Sbjet: 418 FKGITIRSNLLGQSMPISEELWRALELAQAKEPVAALPEQLAPVEAPGRHPSGGQRQR 477

-2391-

Query: 476 LTIARALVQDKIPFLILDDATSALOYLTSARLFKAITKHFWYNLIIVSQRINSIQNADR 535
 L IARAL++ K P LILDDA+SALD T RLFKA+ + + +I+V+Q I ++Q AD+
 Sbjct: 478 LAIARALKPK-PILILDDASSALDNFTKGRFLFKALKKEESDVLVLVTQSTIKNLQFADK 536

Query: 536 ILLLDKRGKQVGFNDHQSLLAHNKVKYKSI 563
 IL+L+G+Q P F +H L N +Y+ +
 Sbjct: 537 ILVLEQGQBLDPASHDQLKVSINALYQEM 564

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2123

A DNA sequence (GBSx2239) was identified in *S.agalactiae* <SEQ ID 6553> which encodes the amino acid sequence <SEQ ID 6554>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 43
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.26 Transmembrane 8 - 24 (1 - 28)
- 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.5904 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- The protein has homology with the following sequences in the GENPEPT database.
- 25 >GP:AAB84433 (GB:AF027868 RAS-related protein [Bacillus subtilis]
 Identities = 53/140 (37%), Positives = 78/140 (54%), Gaps = 2/140 (1%)
- Query: 28 VKKVLQYHDLVQNTLAENGSEANVHLVLSMIYTTETKGDALDVMQSSSESISGTTNSITDSH 87
 ++++ Y LV+ L G L+L M+Y E+K3 D MQSSSES+ N ITD
 30 Sbjct: 49 LERLTDYKPLVEEELSSQGLSNITSLILGMVYQESKKGNDPMQSSSESILGKRNEITDPQ 108
- Query: 88 TSIRKGVITLLSQNISQAKKAKVDVWTVQAYNFGSSYIDYVADHGGENSIELAKNYSKIV 147
 S+K G+ + K+ VD+ T +Q+YN G+ YID+VA+HGG ++ ELAK YS+
 35 Sbjct: 109 LSVKQGKIQFTLMYKTKGKGVLDLTIQSYNMGAGYIDFVAEHGGTHTBELAQYSEQQ 168
- Query: 148 VA--PSLGNVNGDITYFYHP 165
 V P L G+ + +P
 Sbjct: 169 VKINFDLYTCGGNAKNFRYP 188

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4143> which encodes the amino acid sequence <SEQ ID 4144>. Analysis of this protein sequence reveals the following:

- 45 Possible site: 42
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.66 Transmembrane 8 - 24 (7 - 25)
- Final Results -----
 bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 134/200 (67%), Positives = 165/200 (82%), Gaps = 1/200 (0%)
- 55 Query: 1 MFKFLGRLLALIIIIPIGYRLVIIHENVKKVLQYHDLVQNTLAENGSEANVHLVLSMIYT 60
 MF+ LKR + +++ P+ Y+ +IH NV++VL Y +V+ TLAE+ ++ANV LVL+MIYT
 Sbjct: 1 MFRLLKRAKCSFILL-FVLYQSFVIHNVQVRVLAAYKPMVKRTLAENRTKANVDLVLAIIYT 59
- Query: 61 ETKGDAIDVMQSSSESISGTTNSITDSHTSIKRGVITLLSQNISQAKKAKVDVWTVQAYNF 120
 ETKG DVMQSSSES SG NSITDS SI+HGV LLS N++ A++A VD WTVQAYNF

-2392-

Sbjct: 60 ETKGGEADVMSSESSSQKNSITDSQASIEHGVLNLSHNLALAEAGVDSMTAVQAYNF 119
 Query: 121 GSSYIDYVADHGGGNSILARNYSKNVAPSLQNYNGDTPYYHFLALISGGKLYKNGCN 180
 G++YIDY+A+HGG+N+++LA YSK VVAPSLGN +G TYPYHFLALISGGKLYKNGCN
 Sbjct: 120 GTAYIDYIAHHGGQNTVLATYTSKTVVAPSLGNTSGQTYPYHFLALISGGKLYKNGCN 179
 Query: 181 IYYSREVQFNLYLIKIMELF 200
 IYYSREV FNLYLI++M LF
 Sbjct: 180 IYYSREVHFNLYLIKIMELF 199

SEQ ID 6554 (GBS244) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 4; MW 23.1kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 2; MW 48kDa).

GBS244-GST was purified as shown in Figure 211, lane 5.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2124

A DNA sequence (GBSx2240) was identified in *S. agalactiae* <SEQ ID 6555> which encodes the amino acid sequence <SEQ ID 6556>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2401(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9837> which encodes amino acid sequence <SEQ ID 9838> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB71302 GB: A0130879 hypothetical protein [Clostridium
 sticklandii]
 Identities = 32/95 (33%), Positives = 53/95 (55%), Gaps = 1/95 (1%)
 Query: 235 LSPEKLDQQLFDNLTLRLTFVDELKDAIPGFEVQVSDIDHSRQIKLENQKLSLSNGIEL 294
 LS EK + F++ + + L A Q+ + + +K E QK+ +GIB+
 Sbjct: 2 LSEKALETAFRETDKAIYKEALS KAGIENEQI-EVSEETALKRKFBIQKIITEGIEIV 60
 Query: 295 IVPNNVYQDAEVEFEIQNFQDTYSILIKNIQDIQN 329
 +P N Y D +SF+ N DGT S++IKNI +IQ+
 Sbjct: 61 KIPVNYIGPSKLFVANGDGTIVSLVIKNIQNIQS 95

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6557> which encodes the amino acid sequence <SEQ ID 6558>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3336(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2393-

Identities = 246/325 (75%), Positives = 286/325 (87%)

Query: 6 MMDFYIKQIIHQFSPNDTELVDGDTFLTLTPRIDDYFRKKLSKVFSDEAKRGVGEENV 65
 M+D YIK+I+IHQFSPNDTEL+LSD +++TPRID+YFRKKL+KVFSDEAKRGV F +N
 5 Sbjct: 1 MLDYSYIKR+VIRHQFSPNDTEL+LSDSLVSTTPRIDETFRKKLAKVFSDEAKRGVFEANNT 60

Query: 66 FMSHLQDDLYVSSCQIAQLWKSEFVISEDQKTNDLVFTQDKMGHEFAPLRISLKBQFA 125
 F + + DDL +S IAGLWKE FVISEDQKTNDLV+QFEKDG FAFRLI+LKBQFA
 10 Sbjct: 61 FFTTIGDILLSTVIAQLWKSEFVISEDQKTNDLVFTQDKMGHEFAPLRISLKBQFA 120

Query: 126 HVSENGQBPITITTCNNLPSAAQTPEALVFNKSSKQYYLIEKRIHMGSPFANYFSENLIQ 185
 H+S+N E P T+TCNNLPS QTVDEALV+N S QYYLIEKRIHMGSPFANYFSENLI+LL+
 15 Sbjct: 121 HLDNYHEHPTVTTCNNLPSPTQTPDEALVFNKSSKQYYLIEKRIHMGSPFANYFSENLI+LL+ 180

Query: 186 VQFQSVKKSIRKIMBQTAQKIAE+FN+DDF+FGSKMKS ++K LE + LSPKILADQLF 245
 V FQSVKKSIRKIMBQTAQKIAE+FN+DDF+FGSKMKS ++K LE + LSPKILADQLF
 15 Sbjct: 181 VTFQSVKKSIRKIMBQTAQKIAE+FN+DDF+FGSKMKS ++K LE + LSPKILADQLF 240

Query: 246 DDNLTRARILFFVDELDALPGPVQVSDIDHSRQIKKLENQKLSLNGIELIVPMNVQDAS 305
 DDNLTRARILFFVD++FD IP P++SDI+HSRQIKKLENQKLSLNGIELIVPMNVQDAS
 20 Sbjct: 241 DDNLTRARILFFVQKVDVPEPIKSIDHSRQIKKLENQKLSLNGIELIVPMNVQDAS 300

Query: 306 SVEFIQNFQGTYSILIKHIDIQNK 330
 +VEF+ N DGTYSILIKHIDI+K
 25 Sbjct: 301 AVEFLNDDGTYSILIKHIDIQNK 325

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2125

30 A DNA sequence (GBSx2241) was identified in *S. galactiae* <SEQ ID 6559> which encodes the amino acid sequence <SEQ ID 6560>. This protein is predicted to be Serine hydroxymethyltransferase (glyA-1). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3876 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 40

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD35802 GB: AB001743 serine hydroxymethyltransferase [Thermotoga maritima]
 Identities = 243/416 (58%), Positives = 307/416 (73%), Gaps = 7/416 (1%)
 45 Query: 9 KEFDQELWQAIHDEIRQNNIELASRNVSKAVMAQSGVLTNKAEGYPSIRYKGT 68
 K+ D E+++ +E RQ+ +ELIASRNV S RV+ GS+LTNKAEGYPSIRYKGT
 Sbjct: 6 KQVDPEIYELVNLKQKQYQLELIASRNVKAVMAQSGVLTNKAEGYPSIRYKGT 65

Query: 69 DCVDVVESLAIERAKTLFNAEPANVOPHSGSQNNAAAYMALIEPGDTVLGMDLAAGHILT 128
 + VD E AIERAK LF A+ANVOPHSGSQNNAAAYMALIEPGDTVLGMDLAAGHILT
 50 Sbjct: 66 ENVDRAERATERAKRLFCAKFNANVOPHSGSQNNAAAYMALIEPGDTVLGMDLAAGHILT 125

Query: 129 HGASVSPSGKTYHFVSVDFTKTMIDYDNIILKTAQRTQPKLIVAGASAYRIIDFQKR 188
 HGA V+PSGK + V Y V+ +TE +DYD + +A E +PK+IVAG SAY+RIIDF+FR
 55 Sbjct: 126 HGAPVNPSPGKIFKVPYVGNLETETIDYDVRRLALEHKKPIIVAGASAYRIIDFQKR 185

Query: 189 QTADAVDAYLMVDMAHAGLVAAGHHPSPPIYAHVTTTTHKTLRSGRGLILTNDEAIA 248
 +IAD V AYLMVDMAH AGLVA+G HP+P+ YAHV T+THKTLRSGRGLILTNDEAIA
 60 Sbjct: 186 EIADEVGAYLMVDMAHAGLVAAGHHPSPPIYAHVTTTTHKTLRSGRGLILTNDEAIA 245

Query: 249 KKINSVFPGLQGGPLENIVIAKCAVALKEALDESPKTYGSDIINKAQMAKVKEDQDP 308

-2394-

K ++ ++FPG+QGSP L HVIAAKAV KKA+ FK Y + ++KNA+ MA+ F++ +
 Sbjct: 246 KAVDKTI1PFGI1QGSP L MHVIAAKAVCFKRAMTKPEKEYQKQVKNNAKMAKSPFK-RGYR 304
 5 Query: 309 LISDGTDNHFLVDVTVKVIENGKKAQNVLEEVIITLKNISIPFERLSPPFKTSGRIGTFA 368
 ++S GTD HFLVD+T GK A+ LE IT+KNK+IP E+ SPP SGRIIGTFA
 Sbjct: 305 IVSGGDTNHLFLVDLTPKDTIGKAAEKALSCGITT+KNKTIPEKRSPPFVASGIRIGTFA 364
 Query: 369 ITRSGMGVRESRIARLMIKALKN--HMQDVLTEVRQE---IKSITDAFPLYEN 418
 +T+GGM EE IAS++ L N EN V EVR+E ++ L + FFLY +
 10 Sbjct: 365 VITRGKREEMEEIARMIDLVLSNVIDENGTIVKPSRREVSVKRVLECRFPFLTRD 420

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6561> which encodes the amino acid sequence <SEQ ID 6562>. Analysis of this protein sequence reveals the following:

Possible site: 47
 15 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.00 Transmembrane 196 - 212 (196 - 212)
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1001 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GF:CAB15707 GB:Z39122 serine hydroxymethyltransferase [Bacillus subtilis]
 Identities = 250/407 (61%), Positives = 311/407 (76%), Gaps = 2/407 (0%)
 Query: 14 DKELADAIHAEERQSHHIELIASENMVSKAVMAAQGSLVITNKYAGYSGNRYGGTBCV 73
 30 D++++AI E ERQ+ IELIASEN VS+AVM AQGSLVITNKYAGYSGY RYGG E V
 DEQVNAIKNERERQQTIELIASENFVSEAVMEAGQSLVITNKYAGYSGKRYGGGCEHV 67
 Query: 74 DIVETLAIERAKKLFGAAFANVQAHSGSQANAAAYMALIEAGDTVLGMDLAAGHITHGS 133
 D+VE +A +RAK++FGA NVQ HSG+QAN A Y ++S GTVLGM+L+ GSHLTHGS
 35 DVVEDIARDRAKEIFGAESHVNVQPHSGACANMAVITILEQSGITVLGMNLSHSHLTHGS 127
 Query: 134 FVNFGKTYHFGVYSVDITTEMLNYBALEQAKAVQPKLIVAGASAYSRSIDFEKFRJIA 193
 FVNFGS Y+FY Y VD +T+ ++Y+ + E+A A +FKLIVAGASAY R+IDF+KFR IA
 Sbjct: 128 FVNFGVQYNFVEYGVDPKETQYIDYDDVREKALAHKPLIVAGASAYPKTIDFKFRJIA 187
 40 Query: 194 +HVGAYLVMWAHIALVAGVHFSPFPAHYHVTSTHTKTLRSGRGLILINDEALAKKI 253
 D VGRY VMWAHIALVAG+H+FPVFA VT+ITHTKTLRSGS+IL +E KKI
 Sbjct: 188 DEVGAYFVMWAHIALVAGLHPNFVPYADFTVTTHTKTLRSGRGLICREB-FGKKI 246
 45 Query: 254 NSAVFPGQGSPLEHVIAAKAVFKEALDPAPKDVAQALIDTANAAVFAQDORFLIS 313
 + ++FPG+QGSP L HVIAAKAVF E L FK VAQ +I N ++ ++L+G
 Sbjct: 247 DKSIFPGI1QGSP L MHVIAAKAVSFGVLQDPKDTAQNVISNAKRLAEALTKRG-IQLVS 305
 50 Query: 314 GGTENHVFVLVDVTVKVIANGKLAQNLDEVNITLNGIALPFETLSPPFKTSGRIGCAITS 373
 GGTDNH+ LVD+ + GK+A++LDE+ IT NKNIP++ EF TSGIR+G AA+TS
 Sbjct: 306 GGTDNHLLIIVDLRSLGLTGKVAHVILDEIGITSNKNALPIDFSEKFPVTSGRIGTAVTS 365
 Query: 374 RGMVKGESQTIARLITKALVNHIDORTLIEVRQEVRLTDAFELYK 420
 RG + + +I AL NH+ E LEE RQ V LTD FELYK+
 Sbjct: 366 RGFQDGLAEVGAIALNALNHEDEKLEARQVVALTDFELYKE 412

An alignment of the GAS and GBS proteins is shown below.

Identities = 330/417 (79%), Positives = 368/417 (88%)
 60 Query: 1 MIFDKMFKCFDQSLQAIHDEEIRQNNIELIASENVSKAVMAAQGSLVITNKYAGYSP 60
 MIFDK N ++FD+ELN AIH EE RQ+++IELIASEN+VSKAVMAAQGSLVITNKYAGYSP
 Sbjct: 3 MIFDKGNVEDFDKELNDALHAEERQSHHIELIASENMVSKAVMAAQGSLVITNKYAGYSP 62
 Query: 61 SHRYGGTDCVDVVEIASIERAKTLFANFANVQPHSGSQANAAAYMALIEGDTVLGMD 120
 +RYYGOT+CVD+VS+LAIERAK LF A FANVQ HSGSQANAAAYMALIE GTVLGMD

-2395-

Sbjct: 63 GNRYYGTEHCVDIVETLAIERAKKLFGAAZANVQAEHSGSQANAAAYMALIEAGUTVLGMD 122

Query: 121 LPAQGHILTHGASVSFSQKTYHFVSYSVDPKTEMLDYDNLKIAQSTQPKLIVAGASAYS 180
LPAQGHILTHG+ V+FSQKTYHFV YSVD TEMPL+Y+ IL+ A+ QPKLIVAGASAYS 180

5 Sbjct: 123 LPAQGHILTHGSPVNFSGKTYHFVGSYVDTEMLNRYALIEQKAVQPKLIVAGASAYS 182

Query: 181 IIDFSKFRQIADAVDAYLWMDMAHIALGLVAGSHPSPIPIAHVTTTTTHRTLRGPRGGLI 240
IDFEKFR IAD V AYLMVMDMAHIALGLVA+G HFSP+PIAH+ T+THRTLRGPRGGLI 240

10 Sbjct: 183 SIDFEKFRALADHVGAYLWMDMAHIALGLVAAGVHPSFVPIAHVITSTTHRTLRGPRGGLI 242

Query: 241 LTNDLALAKKINSVFPGLQGGPLEHVIAAKAVALKALDPSFKIYGEDIIKNAQANQV 300
LTNDEA+AKKINSVFPGLQGGPLEHVIAAKAVA KEALD+FK Y + II N AMA V 300

Sbjct: 243 LTNDLALAKKINSVFPGLQGGPLEHVIAAKAVAFKEALDPAFKDYAQAIIDNTAAMNAV 302

15 Query: 301 FKEDDDPHLISDGTDNHLFLVDVTKVIENGKKAQNVLEEVNITLNKNSIPFERLSPFKTS 360
F +DD F LIS GTDNH+FLVDVTKVI NGK AQN+L+EVNITLNKN+IPFE LSPFKTS 360

Sbjct: 303 FAQDDRFRLISGGTDNHFVFLVDVTKVIANGKLAQNLIDENVITLNKNAI PFETLSPFKTS 362

20 Query: 361 GIRIGTFAITSRGMGVBSRRIEALIMIKALQNHENQDVLTEVRQETKSLTDAPFLYE 417
GIRIG AITSRGMGV+SS+ IA L+IKAL NH+ +L EVRQEB+ LTDAPFLY+ 417

Sbjct: 363 GIRIGCAITSRGMGVKESQITARLIKALVNHDDRTILSEVRQEBRLTDAPFLYK 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2126

A DNA sequence (GBSx2242) was identified in *S. agalactiae* <SEQ ID 6563> which encodes the amino acid sequence <SEQ ID 6564>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2289 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

35

A related GBS nucleic acid sequence <SEQ ID 9839> which encodes amino acid sequence <SEQ ID 9840> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA035934 GB:AE001752 conserved hypothetical protein [Thermotoga maritima]
Identities = 71/198 (35%), Positives = 114/198 (56%), Gaps = 4/198 (2%)

40 Query: 1 MNDLQQLIEDHGAIVMPETVYGI FAKALSEAVNHYVYLLKRRPKAMNINICDPETIL 60
+ + ++L + +I PTFETVXGI A A +EEA +++LK+RP D + ++I P+ +

45 Sbjct: 17 LKCAEALLKNGEVII PFTETVYGI GADAYNSEACKRI FKLSRKPADNPLIVHISFKQLS 76

Query: 61 KYSNQQPYTLKQLYDAFLGGLITILL-EASQEVFWHNSGLLSVGFPMFPKHPVTLMIAN 119
+ ++ ++L L F P QPLIT+I + S+++P + + L +V RMP HPV L +I

Sbjct: 77 EIAEGYVEPHLDFL-KKFWPGLTIVIPRKKSEKIPPVVTADLPTVAVRMPAHVALKLIEL 135

50 Query: 120 HG-PLIGPSANISCSDCSRGVFSHIGKQFNHGV-LGIEDKALLTGVDSTIIDLSGDRVKIL 177
G P+ PSANISG S + + F +V L I+ G+STI+DL+ ++ +L

Sbjct: 136 FGHPIAAPSANISGRPSATNVKRVIEDFMGKVKLLIIDAGHTPPGLSESTIVDLTKRKFVLL 195

55 Query: 178 RQGAITQEVLTITPEILI 195
R G + R L PEL+

Sbjct: 196 RPSGFVEVERLKLFPPELV 213

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6565> which encodes the amino acid sequence <SEQ ID 6566>. Analysis of this protein sequence reveals the following:

-2396-

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0282 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 127/196 (64%), Positives = 154/196 (77%)

Query: 1 MNDLQILEDHGAVIMPTTETVYIGIFAKALSERAVNHVYELKKRPRDKAMNINICDFETIL 60
 M L I+E A+++PTETVYIG+FAKAL E+AVN VY+LK+RPRDKAMNIN+ DF +IL
 Sbjct: 11 MEYLASIIIESGDALVLPETVYIGLFAKALDEKAVNAVYDLKQRPDKKAMNINVADFNSIL 70

15 Query: 61 KYSINQPTILKQLYDAFLPGPLTIIIEASQEVFINWINGSLLSVGFRMPKHPVTLMIANH 120
 +SK QP YLK+LY AFLPGPLTIIIL+A+ +VF+WINGSL +VGFR+P HP+T +I
 Sbjct: 71 AFSKEQPRYLKLYQAFLLPGPLTIIILKANDQVFINWINGSLTVGFRPLPSHPITAILTQKT 130

20 Query: 121 GPLIGPSANISGCDGSRVFSBIQRFQNHQVLGIEDKALTGVDSTIIDLSGDRVKILRQG 180
 GPLIGPSAN+SG GSRVF I + F+ QV G DD LTG DSTI+DLGG+R IIRQG
 Sbjct: 131 GPLIGPSANLSGKASGRVFDHMQDFDQVFGYADDPFLTGKDSITILDLGSEAVILRQG 190

25 Query: 181 AITQEVLTATIFELIF 196
 AIT+E L A +PEL F
 Sbjct: 191 AITKEILLANVFLRF 205

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2127

A DNA sequence (GBSx2243) was identified in *S. agalactiae* <SEQ ID 6567> which encodes the amino acid sequence <SEQ ID 6568>. This protein is predicted to be protoporphyrinogen oxidase (hemK). Analysis of this protein sequence reveals the following:

Possible site: 50

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07493 GB:AP001519 protoporphyrinogen oxidase [Bacillus halodurans]
 Identities = 94/236 (39%), Positives = 132/236 (55%), Gaps = 12/236 (5%)

45 Query: 49 DTDQQLMENIFQQKKHSP---QYITGKAYFRDLIFVFDVRLVIRPTEBELVDLILSE 105
 + D +L + + + L H S Q++ G F F VD+ VLIRPTEBELV +H E
 Sbjct: 46 ELDGELPQRLEEDLAHAASGVFVCHLIGVBSFYGRQFVQDQHVLIIRPTEBELVLAVLKE 105

50 Query: 106 -----NKVEDCSVLDIGTSGGAIAISLKKERPSQNDVLAISDISALDLAKENANNCDAEV 160
 K E+ ++LDIGTSGGAIA++L E +V A DIS AL +A +NA A V
 Sbjct: 106 IRRQFKKEEITITIDIGTSGGAIAVTLAEEKEEINVTAVDISRDALQVADNARRIGANN 165

55 Query: 161 TFIESDV---FSNISKGFDIIVSNPFYISYNDKDEGVKNVLAESPHSALFADEBGLAIYR 217
 I D+ F +FD+IVSNPFYI +KD + +V EP ALF +GL +YR
 Sbjct: 166 QLHGLDGEPLKTKGTRFDIVISNPFYIPTVEKDTLAVEHVDIEPALAGGVVDGLDYR 225

60 Query: 218 KIENSREYL-QPRGKLYFRIGYKQGGDLRLSLKRYFPNNKRCVLKDFGKDRMVV 272
 +++ + +G + RIG QG D L++ +P VL D D GDR+V+
 Sbjct: 226 RLMSQLPALTKEEGHVALEIGAGQGMVBEKMLQATYKRAVDVLYDENGKDRIVL 281

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6569> which encodes the amino acid sequence <SEQ ID 6570>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4324 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 174/274 (63%), Positives = 207/274 (75%)

Query: 1 MNTAQLIKHYGQLLEACGEEVENFIYVLKDLKQWSTTDYLLNQNSSVSDTDQQLMENIFQ 60
 MNTA LI+ Y LE E+ EN YV +++K+WS+ D L++QN +V+ D L+E+IF
 Sbjct: 1 MNTATLIRTYEDKLEQIDEDRENLA YVFRKIKWSSLDMLHQNCAVTPEDAVLLEHIFC 60

Query: 61 QLKKHRS PQYITGKAYFRDLIFVFDVRVLI PRPTEBELVDLILSENKVEDCSVLIDIGTGS 120
 L +H SPQYITG AYFRDL VD+RVLI PRPTEBELVD+IL+EN +VLDIGTGS
 Sbjct: 61 SISQHLSPQYITGNAYFRDLKI AVDKRVLI PRPTEBELVDMLIARNLDAPLNVLDIGTGS 120

Query: 121 GAIAISLKKERPNDVLASDISVSALDLAKENNNCAEVTFTIESDVFSNLSGKFDTIIVS 180
 GAIAISLKKERP+W V ASDIS +ALDLAK N+ +FTIESDVFS IS FDIIVS
 Sbjct: 121 GAIAISLKKERNMQVTASDISRAALDLAKANADAYQLDITFTIESDVFS LISFTDTIIVS 180

Query: 181 NPFYISYNDKDEVGKIVLASEPHSALFADEGLAIYRKIIENREYLOPRGKLYFBIQYK 240
 NPFYISY DK+EV NVD SEH ALFA E G AIYRKII E + YL GKLYFBIQYK
 Sbjct: 181 NPFYISYEDKEVSLNVQSEPHLAFKENGAIYRKII EQADNYLTKGKLYFBIQYK 240

Query: 241 QGDDRLSLKKRYFPRNRKVLDFGKDRMVVD 274
 Q + ++ +L+ YFF R + DIFGK+RMVV+D
 Sbjct: 241 QABGKMLQAYFPQRIHRAVDIFGKERMVVD 274

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2128

A DNA sequence (GBSx2244) was identified in *S. agalactiae* <SEQ ID 6571> which encodes the amino acid sequence <SEQ ID 6572>. This protein is predicted to be peptide chain release factor RF-1 (prfA).

Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3446 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE15718 GB:Z99122 peptide chain release factor 1 [Bacillus subtilis]
 Identities = 211/351 (60%), Positives = 280/351 (79%), Gaps = 1/351 (0%)

Query: 5 DQLQAVEDRYERLGRLLSDPEVSDTKRPMELSGREASTRETVTAYREYQVQINISDAE 64
 D+L++E+RYE+L ELLSDP+VV+D K+ E S+E++ +ETV YR+Y+ + ++DA+
 Sbjct: 3 DRLLSIEERYEKILNELLSDPEVNDPKKIRYYSKQSDIQETVDVYRQYRDSAEQADAK 62

Query: 65 EMKIDASGDARLEMAKSELKESKAKEEYERLAKILLPKDPNDKNIILSRGAAGGD 124
 M+++ DAE+ +M KEE+ E + E ERLK+LL+PKDPNDKNI+I+IRGAAGG+

-2398-

Sbjct: 63 AMLEEKI-DAEMRDVKEISELQKETETLSERLKVILIPKDPNDQGVIMIRGAAGGE 121

Query: 125 EALFAGDLLIMYQKYARTQGWRFVEMSSVNGVGIGKEVAMVSGQSVSKLYESGAH 184
 EALFAG+L MY +YAE QGW+ EVME++V G GG KE++ M++G YSKLYE+GAH

5 Sbjct: 122 EALFAGNLYRMYSRYEABLQGWKTEVMEVNTGTGYSKEIIFMTISGSAYSILKYENGAAH 181

Query: 185 RVQRVPVPTESQGRVHTSTATVLMPEVEVEVEYIDQKDLRVDIYHAGAGCGQNVNIVATA 244
 RVQRVP TES GR+HTSTATV +FE EEVE +I +KD+KVD + +SG GGG+VN +A

10 Sbjct: 182 RVQRVPETESGGRHTSTATVACLPEAEVEVDIHEKDIRVDTFASSGPGQSVNTMSA 241

Query: 245 VEMVHIPTGIKVEMQERTQQRNRDKAMKIIARAVADHFAQIAQDEQDAERKSTVGTGDR 304
 VR+ H+PTG+ V Q+E++Q KN++KMK++RRA+ D F Q AQ E D RES VQ+GDR

15 Sbjct: 242 VRLTHLPTGVVVSQDEKSGIKNKGAMKVLARAIYDKFQGEAQNETDQTRKSAVSGSDR 301

Query: 305 SERIRTYNFPQNRVTDHRIGLTLQKLDLTLGSKMDEVIDALVMYDQTKLE 355
 SERIRTYNFPQNRVTDHRIGLTLQKLD IL GK+DEV++AL++ DQ KL+

20 Sbjct: 302 SERIRTYNFPQNRVTDHRIGLTLQKLDLTLGSKMDEVIDALVMYDQTKLESLN 352

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6573> which encodes the amino acid sequence <SEQ ID 6574>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3446(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 349/358 (97%), Positives = 354/358 (98%)

Query: 1 MNIYDQLQAVEDRYEKLGEILLSDFDVVSDDTKRFMELSFREASTRETIVTAYREYQVQI 60
 MNIYDQLQAVEDRYEKLGEILLSDFDVVSDDTKRFMELSFRE +TRETIVTAYREYQVQI I

35 Sbjct: 1 MNIYDQLQAVEDRYEKLGEILLSDFDVVSDDTKRFMELSFREETINTRETIVTAYREYQVQIT 60

Query: 61 SDAEMIKDASGDALFEEMAKELKESKAKEERYEERLKIILLPKDPNDQGVIMIRGA 120
 SDAEMIKDASGD ELEEMAKELKESKAKEERYE+LKIILLPKDPNDQGVIMIRGA

40 Sbjct: 61 SDAEMIKDASGDPELEEMAKELKESKAKEERYEERLKIILLPKDPNDQGVIMIRGA 120

Query: 121 AGDEAALFAGDLLIMYQKYARTQGWRFVEMSSVNGVGIGKEVAMVSGQSVSKLYE 180
 AGDEAALFAGDLLIMYQKYARTQGWRFVEMSSVNGVGIGKEVAMVSGQSVSKLYE

45 Sbjct: 121 AGDEAALFAGDLLIMYQKYARTQGWRFVEMSSVNGVGIGKEVAMVSGQSVSKLYE 180

Query: 181 SGARVRQVRVPVPTESQGRVHTSTATVLMPEVEVEVEYIDQKDLRVDIYHAGAGCGQNVN 240
 SGARVRQVRVPVPTESQGRVHTSTATVLMPEVEVEVEY+ID KDLRVDIYHAGAGCGQNVN

50 Sbjct: 181 SGARVRQVRVPVPTESQGRVHTSTATVLMPEVEVEVEYIDPKDLRVDIYHAGAGCGQNVN 240

Query: 241 VATAVRMWHIPTGIKVEMQERTQQRNRDKAMKIIARAVADHFAQIAQDEQDAERKSTV 300
 VATAVRMWHIPTGIKVEMQERTQQRNRDKAMKIIARAVADHFAQIAQDEQDAERKSTV

55 Sbjct: 241 VATAVRMWHIPTGIKVEMQERTQQRNRDKAMKIIARAVADHFAQIAQDEQDAERKSTV 300

Query: 301 TGDRSERIRTYNFPQNRVTDHRIGLTLQKLDLTLGSKMDEVIDALVMYDQTKLESLN 358
 TGDRSERIRTYNFPQNRVTDHRIGLTLQKLDLTLGSKMDEVIDALVMYDQTKLESLN

Sbjct: 301 TGDRSERIRTYNFPQNRVTDHRIGLTLQKLDLTLGSKMDEVIDALVMYDQTKLESLN 358

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2129

A DNA sequence (GBSx2245) was identified in *S. agalactiae* <SEQ ID 6575> which encodes the amino acid sequence <SEQ ID 6576>. This protein is predicted to be thymidine kinase (tdk). Analysis of this protein sequence reveals the following:

5 Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2244 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9841> which encodes amino acid sequence <SEQ ID 9842> was also identified.

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB02289 GB:L0415 thymidine kinase [Streptococcus gordonii]
Identities = 158/189 (83%), Positives = 175/189 (91%)

20 Query: 1 MAQLYYKYGTWNSGKTIILKVAHNYEEQKPVVIMTSALDTRDFGVVSSRIGMRREAV 60
MAQLYYKYGTWNSGKTIILKVAHNYEEQK PVVIMTSA+DTRD G VSSRIGMR+A+
Sbjct: 1 MAQLYYKYGTWNSGKTIILKVAHNYEEQKGVVIMTSADVTRDGVGVSSRIGMKRQAM 60

Query: 61 PISDMDIFSYIQNLFPKPYCVLIDBQFLSKNVYDLARVVDELDPVMAFGLKNDPQN 120
I DD DI YI+NLP+KPYC+LIDE QFL + +VYDLARVVD+LDVPVMAFGLKNDP+N
25 Sbjct: 61 AIEDTDILGYIKNLPEKPYCILIDEAQFLGRHVYDLARVVDELDPVMAFGLKNDPQN 120

Query: 121 NLFEGSKHLLLLADKIDRIKTIQYCSKKA+TWLRTENGKPVYEGDQIQIGNNTYIPVC 180
NLFEGSKHLLLLADKI+EKTIQYCS+KATWVLR+GKPVY+G+QIQIGNNTYIPVC
30 Sbjct: 121 ELFEGSKHLLLLADKIDRIKTIQYCSKKA+TWLRTDHGKPVYEGDQIQIGNNTYIPVC 180

Query: 181 RKHYFPDI 189
RKHYF PDI
35 Sbjct: 181 RKHYFPDI 189

- 35 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6577> which encodes the amino acid sequence <SEQ ID 6578>. Analysis of this protein sequence reveals the following:

Possible site: 39
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2244 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 174/189 (92%), Positives = 184/189 (97%)

Query: 1 MAQLYYKYGTWNSGKTIILKVAHNYEEQKPVVIMTSALDTRDFGVVSSRIGMRREAV 60
+AQLYYKYGTWNSGKTIILKVAHNYEEQKPVVIMTSALDTRD FG+VSSRIGMRREA+
50 Sbjct: 1 MAQLYYKYGTWNSGKTIILKVAHNYEEQKPVVIMTSALDTRDGFVSSRIGMRREA 60

Query: 61 PISDMDIFSYIQNLFPKPYCVLIDBQFLSKNVYDLARVVDELDPVMAFGLKNDPQN 120
PIS+DMDIF++I L +KPYCVLIDE QFLSK+NVYDLARVVD+L+VPVMAFGLKNDPQN
Sbjct: 61 PISDMDIFTTIAQLPEKPYCVLIDBQFLSKNVYDLARVVDELDPVMAFGLKNDPQN 120

55 Query: 121 NLFEGSKHLLLLADKIDRIKTIQYCSKKA+TWLRTENGKPVYEGDQIQIGNNTYIPVC 180
NLFEGSKHLLLLADKIDRIKTIQYCSKKA+TWLRTENGKPVYEGDQIQIGNNTYIPVC
Sbjct: 121 NLFEGSKHLLLLADKIDRIKTIQYCSKKA+TWLRTENGKPVYEGDQIQIGNNTYIPVC 180

-2400-

Query: 181 RKHYFNPDI 189
 RKHYFNPDI
 Sbjct: 181 RKHYFNPDI 189

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2130

A DNA sequence (GBSx2246) was identified in *S.agalactiae* <SEQ ID 6579> which encodes the amino acid sequence <SEQ ID 6580>. Analysis of this protein sequence reveals the following:

10 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3995 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAA26046 GB:M95650 4-oxalocrotonate tautomerase [Plasmid pMW0]
 Identities = 27/60 (45%), Positives = 36/60 (60%)

Query: 1 MPFVKIDLFEGRSQBQKNEAREVTEVVSRIAKAPKNIHVFINDMPESTYYPQGLKKK 60
 MP +I + EGRS EQK L REV+E +SR AP ++ V I +M +G + GEL K
 Sbjct: 1 MPFAQIHILFGRSDBQKETLIREVSEAFSLDAPLTSVRVIITEMAKGHFPIGGLASK 60

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6581> which encodes the amino acid sequence <SEQ ID 6582>. Analysis of this protein sequence reveals the following:

30 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.4128 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 56/60 (93%), Positives = 59/60 (98%)

40 Query: 1 MPFVKIDLFEGRSQBQKNEAREVTEVVSRIAKAPKNIHVFINDMPESTYYPQGLKKK 60
 MPFV IDLFEGRSQBQKRN-LAREVTEVVSRIAKAPKNIHVFINDMPESTYYPQGR+K+K
 Sbjct: 1 MPFVTIDLFEGRSQBQKQLAREVTEVVSRIAKAPKNIHVFINDMPESTYYPQGRKQK 60

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 2131

A DNA sequence (GBSx2247) was identified in *S.agalactiae* <SEQ ID 6583> which encodes the amino acid sequence <SEQ ID 6584>. Analysis of this protein sequence reveals the following:

50 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2154 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9843> which encodes amino acid sequence <SEQ ID 9844> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP: AAC65759 GB: AB001250 conserved hypothetical protein [Treponema
    pallidum]
    Identities = 103/317 (32%), Positives = 163/317 (50%), Gaps = 15/317 (4%)

10 Query: 7  QLSHSLRMGTITDIQINSKNAQKQIR---EVIELLELYKNRFSANDFNSRLMAINNNA 62
    + S + ++GT +++ SK ++ EV LL+ + SAN +S L A+N A
    Sbjct: 31 EYSRAELVIGTLGRVVRVYSKRPAAEVHAALREVFTLLQQQEMVLNSNRDDSLAALNAQA 90

    Query: 63 GIKPIQVHFDLFLITIGKEHSLARPSNLNIAIGPLVQTVWRIGFSDAKLPSPSEISEAMI 122
    G P+ V L+ L+ + N A+G V+ W IGF A+ P P + EA+
15 Sbjct: 91 GSAPVVVDRSLYALLERALLFAEKSOGAPFNALGAXVKLWNIIGPDRAVDPDAKALKALT 150

    Query: 123 LSDPTHILDSN-----KQSVFLNIQGMKIDIGALAGYIADKIMTYLKNEMDSAINL 177
    D + L + +V L Q GM++DLGA+AKG++ADKI+ L +DSA+++L
20 Sbjct: 151 RCDFRQVHLRAGVSVGAPHTVQLAQAGMQLDLGALAGKFLADKIVQLLTAHALDSALVDL 210

    Query: 178 GGNV----LVHGDNPNRSEGY--WVIGIQHPKKRGNIGTVKIKNQSVVTSGYERRLI 231
    GGN+ L +GD + + W +GI+ P K V +++ SVVTG YER
    Sbjct: 211 GGNIFALGLKYGDVRSAQAQLNWNVIRDPHSTGQKPAVSVVRDCSVTSGAYRFFE 270

25 Query: 232 IDKEYYHIFDRQTGYPIQTEMASISIVSKQSVDCIEWTLRFLGLSKEALDIILNAVSYI 291
    D YHI D TG+P T+ S+SI +S D + T F L ++ +L +
    Sbjct: 271 RCGVRYHILDPVTGFPAHTDVDSVSIFAPRSTADALATACFVLGYEKSALLRFFPGV 330

30 Query: 292 EGIITTKDDRIYLSGL 308
    + + I D R+ S G+
    Sbjct: 331 DALFFFDKRVRSAGI 347
  
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6585> which encodes the amino acid sequence <SEQ ID 6586>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 52
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.1020 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

An alignment of the GAS and GBS proteins is shown below.

```

45 Identities = 182/310 (58%), Positives = 232/310 (74%)

    Query: 8  LSHSLRMGTITDIQINSKNAQKQIREVIELLELYKNRFSANDFNSRLMAINNNAQIKPI 67
    ++ L+LMGT IDIQI S A +Q+ VI+LL YKNRFSAND NSELMATINNAQV+KP+
    Sbjct: 3  VTQQLKLMGTVIDIQIESDKACQQLSRVIDLLYTYKNRFSANDSSELMATINNAQV+KPV 62

50 Query: 68  QVHFDLFLITIGKEHSLARPSNLNIAIGPLVQTVWRIGFSDAKLPSPSEISEAMISDPT 127
    VH DLF LI TGR RSL+ PSNLNIAIGPLVQ WRIGF DA++PS + IS+ +L+DP
    Sbjct: 63 SVHSDFNLITIGKHSLSSTPSNLNIAIGPLVQAWRIGFEDARVPSNLISQQLALTDP 122

55 Query: 128 HILLDSNKQSVFLAQGMKIDIGALAGYIADKIMTYLKNEMDSAINLGNVLVHGDN 187
    +L+D KQ+VFL Q+GM +DLGALAGYI DKIM YL + IDSA+INLGNV VHG N
    Sbjct: 123 QVLIDDKKQTVFLQQVGMALDLGALAGYITDKIMAYLIEDGIDALINLGNVRVHGPV 182

60 Query: 188 PNRSEGYWVIGIQHPKKRGNIGTVKIKNQSVVTSGYERRLIIDKEYYHIFDRQTGY 247
    P + + IGIQ P KRG+++G +K+ N SVVTG YER+ K+YHII DRQTY
    Sbjct: 183 PKSPDKTFRIGIQKFDKRGQHLGVKVNHSVVTSGITYERQFTSGKCYHILDKQTYG 242

    Query: 248 PIQTEMASISIVSKQSVDCIEWTLRFLGLSKEALDIILNAVSYIEGIITTKDDRIYLSG 307
  
```

-2402-

Pi+T+M S++I++ S C+IWTRRLGL + +LN IBG+++T+ + +S+G
 Sbjct: 243 PIETDMLSLTINAPSSFYCDIWTIRLGLDSSMITILMTFONIEGLLVTRKHVILMSNG 302

Query: 308 LKHHPQLFPH 317
 L+H+PQ +YH
 Sbjct: 303 LRHYFQPYH 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 2132

A DNA sequence (GBSx2248) was identified in *S. agalactiae* <SEQ ID 6587> which encodes the amino acid sequence <SEQ ID 6588>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0966 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:ANG18632 GB:AY007504 unknown [Streptococcus mitis]
 Identities = 92/160 (57%), Positives = 119/160 (73%), Gaps = 1/160 (0%)

Query: 1 MKLIGIVGINSNKSTNRQLLQVNOQHPADKAEIELIEVDLPLNPKFADINVPQVILIDIA 60
 NKL+ IVGTNSN+STNR+LL++NQ+HF+DKA+IE++E+K LP FN+P D+ P + +
 Sbjct: 1 MKLVAIVGTNSNRSTNRKLLKFPQKHFSKADIEVLEIKQLPAFNEPEDEQAPAEVQAFA 60

Query: 61 AKIETDGVIIOTPEYDHSIPSALMSVLAWLSYGIYPLANKPMVTGASYGTLGSSRAQL 120
 KI DGVII TPEYDH+IP+ L S L W++Y L+NKP MI GAS G LG+SRQA
 Sbjct: 61 EKILAADGVIIISTPEYDHTIPAPLASALEWIAYTSRALINKPTMIVGASIGLLGTSRAQA 120

Query: 121 QLRQILNAPELKASVLP-DEFLSHSLQAFKDGKGNLHDIE 159
 LRQIL+APELKA V+P EF L HS Q D + +L+ E
 Sbjct: 121 HLKQILDAPELKARVMPGTTEFFLGHSSEQLDDECHLHFE 160

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6589> which encodes the amino acid sequence <SEQ ID 6590>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB62679 GB:AL133422 putative secreted protein. [Streptomyces coelicolor A3 (2)]
 Identities = 68/192 (35%), Positives = 94/192 (48%), Gaps = 25/192 (13%)

Query: 4 ILPIVGSRLREGSFNHQLAAQAQK-ALEHQAVVSYLAWKDVVFNQDIEANAPLPVDA-- 60
 IL +VGSRL GS N QLA A + A E V + ++P N+DI+ +P A
 Sbjct: 5 ILALVGSRLAGSHNRQLAEAAVRAFPAGARVQLFEGLAETPPYNEIDIVDGSVPAAAKL 64

Query: 61 RQAVQSADAIWIFTVPYVNFSPGSVKNLQMLSRALDLSDPGPSAIGKVKVTVSSVANG 120
 R+A Q A A +F+P YN +IP +KN +DWLSR P G A GK V V A G
 Sbjct: 65 REAAQGAQAFLFPSPEYNGTIPAVLKNAIDWLSR-----PYGAGAITGKPVAVGTAFG 118

-2403-

Query: 121 GHDOVFQDQFA-----LLPFIKTSVAGSEFTK-ATVNP--DAWGTRLEISKETKA 167
 + V+ Q +A ++ I+ S+ G T+ A +P DA +L E A
 Sbjct: 119 QYGGVWQDSEARKAVGIAGGKVTIEDIKLSIFGSVTRPAETHPADDAEVAQAQL--TEVVA 175

Query: 168 NLLSQAEALLAA 179
 L A+ +AA
 Sbjct: 176 RLHGHADAETAA 187

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 28/90 (31%), Positives = 49/90 (54%)

Query: 3 LIQIVGTNSNKSINQQLQYMQQHFADKAKIELIEVKDLPLFNKPADKNVPQVILDIAAK 62
 ++ IVG+ S N CL Q+ +A + + KD+P+ N+ + N P ++D
 Sbjct: 4 ILPIVGLSREGSPNQIAAQAQKALEHCAVSVYLNWKKDVPVLANODIEANAPLEVVDAQQA 63

Query: 63 IEETDGVIIQTPFYDHSIPSALMSVLAWLS 92
 ++ D + I TP Y+ SIP ++ ++L WLS
 Sbjct: 64 VOSADAINIFTFVYNFSIPGSVKNLLDWLS 93

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2133

A DNA sequence (GBSx2249) was identified in *S. agalactiae* <SEQ ID 6591> which encodes the amino acid sequence <SEQ ID 6592>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1160 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2134

A DNA sequence (GBSx2250) was identified in *S. agalactiae* <SEQ ID 6593> which encodes the amino acid sequence <SEQ ID 6594>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2132 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:ANG18632 GB:AY007504 unknown [Streptococcus mitis]
 Identities = 80/162 (49%), Positives = 112/162 (68%)

Query: 1 MKFVGIVGSNAEQYNRMLLRFIRKNFTKFELEVLIDDIPMPNQDNWESPOLRLN 60

-2404-

MK V IVG+N+ +S MR LL+F++K+F K ++EVLRI +P FN+ ++ + +++ +
 Subjct: 1 MKLVAIVGTSNSRSTRRLKLFQKHFSKADIEVLRIQLPAPNSPDEQAPAEVQAPS 60
 Query: 61 NKITRADGVIIATPERNHITTAALKSVLEWLSFAVHPELENKPMIVGASYYDQTSRAQI 120
 KI ADGVII+TPE+HTI'A L S LEW+++ L NKP MIVGAS GTSRAQ
 Subjct: 61 EKILAADGVIIISTPEYDHTIPAPLASALEWIIATYTERALINKPTMIVGASLLOLTSRAQA 120
 Query: 121 HLRKILDAPGVNAYTLPGNEFLIGKAKEAFDONGNIINPGTV 162
 HLR+ILDAP + A +PG EF LG +++ DD ++ NF V
 Subjct: 121 HLRQILDAPELKAERVMPGTETFFLGHSBQVLDDBCILANNFEKV 162

There is also homology to SEQ ID 6596.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 2135

A DNA sequence (GBSx2251) was identified in *S.agalactiae* <SEQ ID 6597> which encodes the amino acid sequence <SEQ ID 6598>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 20 INTEGRAL Likelihood = -7.32 Transmembrane 13 - 29 (11 - 29)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3930 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2136

A DNA sequence (GBSx2252) was identified in *S.agalactiae* <SEQ ID 6599> which encodes the amino acid sequence <SEQ ID 6600>. This protein is predicted to be potential nitrite transporter. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 35 INTEGRAL Likelihood = -9.92 Transmembrane 61 - 77 (54 - 82)
 INTEGRAL Likelihood = -5.57 Transmembrane 106 - 122 (103 - 126)
 40 INTEGRAL Likelihood = -5.15 Transmembrane 160 - 176 (159 - 177)
 INTEGRAL Likelihood = -4.09 Transmembrane 180 - 196 (179 - 199)
 INTEGRAL Likelihood = -1.01 Transmembrane 233 - 249 (233 - 249)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15832 GB:Z99123 alternate gene name: ipa-48r-similar to
 50 nitrite transporter [Bacillus subtilis]
 Identities = 82/253 (32%), Positives = 119/253 (46%), Gaps = 10/253 (3%)
 Query: 6 EKIAYNCAKEALYKESLRVYALRSLAGAYLIWSTAGIIVAADTIGK-ISPALSQFVF- 63

-2405-

+K+ KK+ ++ S RY LRS+LA ++ GI AA G A S F F
 Sbjct: 7 QRVEQYALKKQNI PASSKIRYVLRSLIASIFIGF---GITAASKGSGYFFMADSPFAPF 62
 Query: 64 --AFIFSPGLIYVLIFNGELATSNMMLYLTAAYNNISWKKAMTILYCTFFNLVAGCIL 121
 A F ++ + G+L T N Y T A K ISW+ ++ + N+GA +
 Sbjct: 63 ARAVTFGAAILMAYGGGDLFTGNTFYPTTALRCKISWRDITLYLMSSSYAGNLIGAILF 122
 Query: 122 AMLFNQSYSPQMLTNDSPFLSHVAVAKKLGKPSGAPLEGIIANNFVNLAIALAYMLKRESS 181
 A L + + F + SFL H+ K+ P+S F G+ + N V LA M LK E A
 Sbjct: 123 AILISATGLFEPSVHSFLIHLAEHKMEPPASSELFFRGMLCNWLVCLAFFIPMSLKGGGA 182
 Query: 182 KMTVLSAIFMFPVLSNEHLIANFASPMIAAFAFHHIRKGPILLNIIRQWTLVFFGNWIG 241
 K+ ++ +F F EH IAN +F ++ IEH TL+ +R V GN
 Sbjct: 183 KLFITMLFVPCFIFSGFHSIANMCTFAISLL--IEHDTVTLMAVRNLIPVLGNLTA 240
 Query: 242 GGVFIGLAYAWLN 254
 G V +G Y LN
 Sbjct: 241 GIVNNGMYITLN 253

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6601> which encodes the amino acid sequence <SEQ ID 6602>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -9.77 Transmembrane 142 - 158 (139 - 171)
 INTEGRAL Likelihood = -9.34 Transmembrane 95 - 111 (89 - 119)
 INTEGRAL Likelihood = -2.02 Transmembrane 61 - 77 (61 - 79)
 INTEGRAL Likelihood = -1.12 Transmembrane 261 - 277 (261 - 279)
 INTEGRAL Likelihood = -0.53 Transmembrane 191 - 207 (191 - 207)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has homology with the following sequences in the databases:

>GP:AA80864 GB:U93874 foxmate dehydrogenase [Bacillus subtilis]
 Identities = 133/258 (51%), Positives = 181/258 (69%)
 40 Query: 36 KTPQILQIAIHIGHKYKTKFLAKSILGFIGGAMISLGILLYVRIAASGLETFGAPSSI 95
 + P+I EA I G K+ + +LGF+GGA I+LGILL +R+ + +G+ SS+
 Sbjct: 4 RKPEDEAIAEAGKIKLFLSLVLGLGGAFLAIGYLLDIRVIGDLPKFWGSLSSSL 63
 Query: 96 VGACAFPIGLIIIIHAGGELITGNMMVAASALLAKCIKPSELAKMLITLNFVIGAVFV 155
 +GA FF+GLI++++AG ELITGNMM+V+ AL +KI ELA NW I+T+ N+IG+FF
 45 Sbjct: 64 IGRVFPVGLILVLAGELITGNMMVSMALPSKIKSVKELAINWIGVITNINLIGALFV 123
 Query: 156 AFVFGHFLGLTSGAGIFKEVIEVAHAKIAASPLQALVSGISGNMFVGLALMLCYGNDRA 215
 A+ FGH +GLT G + E+ I VA K+ S + L+S IGCNW V LA+WL +GA DRA
 50 Sbjct: 124 AYFFGHGLVGLTGTGAYLEKTIIVAQGLDMSFGKVLIAIGQNLVCLAVWLSFGQNDRA 183
 Query: 216 GKFLGTWFPVMTFVALGPOHSVANAIPVPAIFEGGATWLDVFTNTFIVVSIGNIIGQAF 275
 GK LG NFP+M FVA+GQH VAN FVIEAIF G TW F+ N I + GH+IGGA+P
 Sbjct: 184 GKILGINFPIMAFVAGFQHVANMFVPAIFAGSGTQQFIGNIIPAFIGNVIGGAFF 243
 55 Query: 276 VSPFLYFKVYYHPOKSKTQ 293
 V +YF Y+ +S+ +
 Sbjct: 244 VGLIYFIAYHKKORSKE 261

An alignment of the GAS and GBS proteins is shown below.

60 Identities = 69/240 (28%), Positives = 101/240 (41%), Gaps = 18/240 (7%)
 Query: 15 KEALYKESIGRYALRSMLAGYLTNSTAAGIVAAITGKISPAISGQFVAFIFSPGLIYV 74
 K L K LG + G L + AA +T G A S V A F GLI +
 Sbjct: 55 KTFIAKSLGFIGGAMISLGILLYVRIAAS--GLETFG---AFSSIIGACAFPIGLIII 108

-2406-

Query: 75 LIFNGELATSNWLYLTAGAYNNKNISWKKAMTILYCTFFNLVGACILAWIFNQSYSPQL 134
 L+ GEL T NM+ +A K I + + + T FN++GA +A++F F L
 Sbjct: 105 LMAGGELITGRWMAVSAALLAKKIKPSLAKNMLITLLENVIGAVFVAFVGH---FLGL 165

Query: 135 TNDSPFLGHVAK---KLKGPSSGAFLEGITIANMFVNLAIIAYMLLKEESAMTIVLSAI 190
 T+ V + K+ A + GI N FV LA+ + + K +
 Sbjct: 166 TSAGIFKEEVIEVAHAKIAASPLQALVSGICNWFVGIALWLCYGANDAAGKFLGTWFPV 225

Query: 191 PMFVFLSNEHLIANFASFMIAAPSHIHIGFTLINIRQWTLVFPFGNWIGGVFLGLAY 250
 FV L +H +AN A F G T L+ + + V+ GN IGG +F+ Y
 Sbjct: 226 NTFVALGFQHSVANAFVIPAALFE-----GGATWLDFTVNFIFVYSNTIGGALFVSFLY 280

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2137

A DNA sequence (GBSx2253) was identified in *S.agalactiae* <SEQ ID 6603> which encodes the amino acid sequence <SEQ ID 6604>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1342(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2138

A DNA sequence (GBSx2254) was identified in *S.agalactiae* <SEQ ID 6605> which encodes the amino acid sequence <SEQ ID 6606>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.22 Transmembrane 44 - 60 (44 - 60)

----- Final Results -----
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2407-

Example 2139

A DNA sequence (GBSx2255) was identified in *S.agalactiae* <SEQ ID 6607> which encodes the amino acid sequence <SEQ ID 6608>. This protein is predicted to be xanthine permease (pbuX). Analysis of this protein sequence reveals the following:

```

5  Possible site: 23
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.91    Transmembrane 160 - 176 ( 156 - 188)
      INTEGRAL    Likelihood = -6.48    Transmembrane 184 - 200 ( 179 - 211)
      INTEGRAL    Likelihood = -6.21    Transmembrane 101 - 117 ( 96 - 121)
10  INTEGRAL    Likelihood = -4.04    Transmembrane 309 - 325 ( 306 - 332)
      INTEGRAL    Likelihood = -3.98    Transmembrane 334 - 350 ( 331 - 353)
      INTEGRAL    Likelihood = -3.88    Transmembrane 400 - 416 ( 396 - 420)
      INTEGRAL    Likelihood = -3.45    Transmembrane 19 - 35 ( 18 - 38)
      INTEGRAL    Likelihood = -2.81    Transmembrane 127 - 143 ( 127 - 146)
15  INTEGRAL    Likelihood = -2.71    Transmembrane 228 - 244 ( 227 - 249)
      INTEGRAL    Likelihood = -2.02    Transmembrane 47 - 63 ( 47 - 63)
      INTEGRAL    Likelihood = -1.97    Transmembrane 75 - 91 ( 73 - 92)
      INTEGRAL    Likelihood = -0.85    Transmembrane 368 - 384 ( 368 - 384)

20  ----- Final Results -----
      bacterial membrane --- Certainty=0.4163 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25  The protein has homology with the following sequences in the GENPEPT database.
   >GP: CAB14123 GB: Z99115 xanthine permease [Bacillus subtilis]
   Identities = 213/412 (51%), Positives = 292/412 (70%), Gaps = 5/412 (1%)

30  Query: 14  LGLQHLLAVYAGSILVPIMIASALGVNAKQLTYLTIADIPMCGIATLQLRLSKHFVGVL 73
      LG+QH+LAVYAG+I+V+++ A+G +QLTYL+ DIPMCG+ATLLQ+ ++ FG+GL
      Sbjct: 11  LGLQHVLAVYAGAIIVPLIVGKAMGLTVBQLTILNSIDI PMCGVATLLQVSNRFFSIGL 70

   Query: 74  FVVLGCAFGSVAPLSIGAQQSGTYMFGALIASGIYVVLVAGIPSKVANFFPPPIVTSVSI 133
      FVVLGC F +V+P+ IG++ G ++G++IASGI V+L++ F K+ +FFPP+VTGSV+
35  Sbjct: 71  FVVLGCTPTAVSPMIAGSEYGVSTVYGSIIASGILVILISFFPGKLVSFPPVVTGSVV 130

   Query: 134  TTIGLTLIPVAMGNMD---NAKESLQSLTSLVLTIGVLLINIFAKGLKLSISILIGL 190
      T IG+TL+PVAM NM +A L +L L+ + +++L+ F KGF+KS+SLIG+
40  Sbjct: 131  TTIGTILMPVAMNNHAGEGASDFGLDLSHLALAPTLVLSIIVLLYRPTGFIKSVSLIGI 190

   Query: 191  ISGTILAAFMGLVDAASVADAPLVIKPPFYFGAPRFEFTSLIMMCIATVSNVESTGVY 250
      + GT +A FMG V V+DA +V + +PFYFGAP F I+ M I+A VS+VESTGVY
45  Sbjct: 191  LGTFIAYPMGKGVQFNDVSNAAVQCMQPFYFGAPSPHAPIITMSIVAIIVSLVESTGVY 250

   Query: 251  LAISDITNDKLDKSRKNGYRSEGLAVILGGLPNTFFPYTGFQNVGLQISGIRKRPYI 310
      AL D+TN +L L GYR+BGLAVILGG+FN PPT FSNQVGLVQ++GI+ I
50  Sbjct: 251  FALGDLTNRLITEDLSKGYRABGLAVILGGLPNAFPYTFATFQNVGLQITGIRKRAVY 310

   Query: 311  FTALFLVILGLLPKFGMAQMIISPVLGGAMVLFGMVALQGMKMLNVDPERNEHPII 370
      T + L+ GL PK A +IPS VLGGAM+ +FGMV G+KML++DP E N +I
55  Sbjct: 311  VTGVILMAFLGFLPKIAAFTTIIISAVLGGAMVAFGMVIAVYGIKMLSRIDFAKOE-NLI 369

   Query: 371  AAVSIAGVGPNGT-NLFSISLNTLQMFITNGIVISTLTAVVLNIILNGLPK 421
      A S+ G+G ++F LP+ L + TNGIV + TAVVLNI+ N K
      Sbjct: 370  VACSVGLIGLVTVVVDPIPKQLPSAULTLTNGIVAGSPTAVVLNIIVYNVSK 421

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6609> which encodes the amino acid sequence <SEQ ID 6610>. Analysis of this protein sequence reveals the following:

```

60  Possible site: 29
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.32    Transmembrane 160 - 176 ( 158 - 181)
      INTEGRAL    Likelihood = -6.37    Transmembrane 103 - 119 ( 98 - 124)

```

-2408-

- INTEGRAL Likelihood = -5.84 Transmembrane 130 - 146 (126 - 152)
 INTEGRAL Likelihood = -5.68 Transmembrane 187 - 203 (182 - 207)
 INTEGRAL Likelihood = -3.98 Transmembrane 337 - 353 (334 - 356)
 INTEGRAL Likelihood = -3.82 Transmembrane 232 - 248 (225 - 252)
 INTEGRAL Likelihood = -3.35 Transmembrane 403 - 419 (399 - 421)
 INTEGRAL Likelihood = -2.50 Transmembrane 22 - 38 (21 - 41)
 INTEGRAL Likelihood = -2.07 Transmembrane 312 - 328 (312 - 328)
 INTEGRAL Likelihood = -1.97 Transmembrane 78 - 94 (76 - 95)
- 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
- 15 The protein has homology with the following sequences in the databases:
 >GP:CAB15234 GB:Z99120 similar to purine permease [Bacillus subtilis]
 Identities = 216/421 (51%), Positives = 302/421 (71%), Gaps = 5/421 (1%)
- 20 Query: 6 KQEHSHSQSAVLGLQHVLMSNTAGSILVPIIMIALGALYSARELTLYLSTIDIPMGQVATFLQ 65
 K++H+ Q +LGLQH+L+MIAG+ILVP+++ A+G +A +LTYLI+ D+FMCG+AT LQ
 Sbjct: 2 KEQENALQLNMLGLQHMLAMTAGAILVPLIVGAAILGAGQLTYLITAILDIPMGQATFLQ 61
- 25 Query: 66 LKLTGHTGVGLFVVIGCAPSQSVAPLSIIQAQQSGMFGALIASGIYVVLVAGIFSKIAR 125
 L ++ G+GLFVVVLGC F +V P+ IG+ G A++GA+LH+G+ V+L AG F K+ R
 Sbjct: 62 LWRNRIFYGIGLFVVVLGCTFTAVGPMISIGSTGYVEAITGAILIAGLIVLVAAGFFGKLVR 121
- 30 Query: 126 FFFPIVGTGSVITVIGLSLVGVAMGM--GENVKE-PTAQSMMLSLTITVILLVQKPTKG 182
 FFFP+VTGSV+ +IG+SL+ AM N+ G+ KE + +++L ILL+ F KG
 Sbjct: 122 FFFPVTGSVMVMIIGLSIPTAMNLAGGSGSKFQFSLONVLLGFGVTAFILLLYFFFG 181
- 35 Query: 183 FVKSISILIGLVAGTLVSAMMGLVDITPVVEASMIHVPTFFYFGNPTFEITSIVMMCIA 242
 F++SI+IL+GL+AGT + MG VD + V+EAS+HVF+ FIFG PTFE+ ++V M +A+
 Sbjct: 182 FIRSIALILGLLGTAAATFMKGIVDSFVLEASMLHVPSLFYFGPTFELPAVVMILVA 241
- 40 Query: 243 TVSMVSTGVYLLASDLINDQLDEKRLNGYRSBGIAVLAGLNFNTFFYTFGSGNVGLVQ 302
 VS+VSTGVY AL+D+IN +L EK L GYR+EG+A+ LGGLEN FYTF FSGNVG+VQ
 Sbjct: 242 IVSLVSTGVYFALADITNRRLSEKLEKYSRABGLAILLGLGNAPFFYTFPSQNVGIVQ 301
- 45 Query: 303 ISGIKTRRPYYAAGILVIGLLPKFRAMQNPSPVLSGAMLVFGMVALQSGMGLNRV 362
 +S +K+ I ILV IGL+PK A+ +IP+VVLGSM+V+FGMV G++ML+ L
 Sbjct: 302 LSVKMSVNVIAITGIIILVAILGLVPKAAALITVITPVVLSGAMIVFGMVISYGIKMLSV 361
- Query: 363 DFQNEYNFIAAVSIAGLGFNQT-NLFAFLPETQMFNTNGIVIAITLSVVLALVNGK 422
 D ++ N +I A S+S GLG LF+SL A + +GIIV +LT++ L+ K
 Sbjct: 362 DLD-SQCNLLI IASSVSLGATVPAFLSSLSGASVLAGSGI VIGSLTALIAHAFQTK 421

An alignment of the GAS and GBS proteins is shown below.

- Identities = 328/416 (78%), Positives = 380/416 (90%)
- 50 Query: 7 SNSQAALLGLQHLLAMYAGSILVPIIMIASALGYNAQLTYLITIDIPMGQIATLLQLRLS 66
 S+SQ+A+LGLQH+L+MYAGSILVPIIMIA ALGY+A+LTYLI+TDIPMGQ+AT LQ+L+
 Sbjct: 10 SESQSAVLGLQHVLMSNTAGSILVPIIMIALGALYSARELTLYLSTIDIPMGQVATFLQKLT 69
- 55 Query: 67 KHFGVGLFVVVLGCAFSQSVAPLSIIQAQQSGMFGALIASGIYVVLVAGIFSKVNVFFPP 126
 KH GVGLFVVVLGCAFSQSVAPLSIIQAQQSG MFGALIASGIYV+LVAGIFSK+A FFFP
 Sbjct: 70 KHFGVGLFVVVLGCAFSQSVAPLSIIQAQQSGMFGALIASGIYVVLVAGIFSKIARFFPP 129
- 60 Query: 127 IYVSGVITITGLTIPVAMGMNDNAKPSLSQSLTSLVITGVVLLINIFAKGLKSIISI 186
 IYVSGVIT IGL+L+ VAMGMNDN KSP+ QS+ LSL+TI +LL+ F KGF+KSIISI
 Sbjct: 130 IYVSGVITVIGLSLVGVAMGMNDNVKRP+TAQSMMLSLTITVILLVQKPTKGFKVKSISI 189
- 65 Query: 187 LIGLISGTILAAFMGLVDASVADALVHIIKPPFYGPAPRPFETSILAMCIIATVSMSVS 246
 LIGL++GT++A MGLVD + V +A +H+P FFYFG P FE TSI+MCCIATVSMSVS
 Sbjct: 190 LIGLVAGTLVSAMMGLVDITPVVEASMIHVPTFFYFGNPTFEITSIVMMCIIATVSMSVS 249
- Query: 247 TGVYIALSDITNDKLDKSKRLNGYRSBGLAVILGGLNFNTFFYTFGSGNVGLVQISGRTR 306

-2409-

TGVVIALSD+TND+LD KRLRNGYRSEGA+V LGGLPNTFFPTGFSQNVGLVQISGI+TR
 Sbjct: 250 TGVVIALSDLTNDQDLEKRLRNGYRSEGAIVFLGGLPNTFFPTGFSQNVGLVQISGIKTR 309
 Query: 307 KPIYFTALFLVITIGLLPKFGMAQNI PSVLGGAMLVLPGMVALQGMQLAKVDFEHEH 366
 +PIY+ A LV++GLLPKF AMQNI PSVLGGAMLVLPGMVALQGM+MLN+VDF+ NE+
 Sbjct: 310 RPIYAAAILVITIGLLPKFRMAQNI PSVLGGAMLVLPGMVALQGMQLAKVDFQINEY 369
 Query: 367 NFIIAAVSIAGVGFGKTNLFISLDPNLTQMFLTNGIVISTLATVVLNIIILNGLPKK 422
 NFIIAAVSI+AG+GFGKTNLF SLP T QMFLTNGIVI+YLT+VVLN++LNG K+
 Sbjct: 370 NFIIAAVSIAGLFGKTNLFASLPETAQMFLTNGIVIAITLTSVVINLVINGKDKQ 425

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2140

- 15 A DNA sequence (GBSx2256) was identified in *S. agalactiae* <SEQ ID 6611> which encodes the amino acid sequence <SEQ ID 6612>. This protein is predicted to be xanthine phosphoribosyltransferase (xpt). Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1921 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 The protein has homology with the following sequences in the GENPEPT database.
 >GP:CAAL3587 GB:AJ233894 xanthine phosphoribosyltransferase
 [Streptococcus pneumoniae]
 Identities = 133/162 (82%), Positives = 144/162 (88%)
 Query: 16 GENILKVDSPFLTHQVDFELMQEIGKVPADKYKEAGITKVVTIEASGIAPAVTAQAQGVF 75
 G+NLKVDSPFLTHQVDF LM+HIGKVPFA+K+ AGITKVVTIEASGIAPFA+ A+AL VP
 Sbjct: 1 GDNILKVDSPFLTHQVDFSLMRBIGKVPFAEKFPASAGITKVVTIEASGIAPALPTAEALNVP 60
 Query: 76 MIPAKKAKNITMTGILTAEVYSPTKQVTSQVSVISRFSLNDVTLLIIDDFLANGQAAGK 135
 MIPAKKAKNITM EGILTAEVYSPTKQVTS VSI +FLS +D VLIIDDFLANGQAAGK
 Sbjct: 61 MIPAKKAKNITMTGILTAEVYSPTKQVTSVSIAGKFLSPEDKVLIIIDDFLANGQAAGK 120
 Query: 136 LLEIIQAGAKVAGIGIVIEKSPQGRDLLEKTVGPVTSLAR 177
 L++II QAGA V IGIIVIEKSPQGRDLLEK G PV SLAR
 Sbjct: 121 LIQIIQAGATVEAIGIVIEKSPQGRDLLEKAGVPLSLAR 162

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6613> which encodes the amino acid sequence <SEQ ID 6614>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2576 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 156/193 (80%), Positives = 172/193 (88%)
 Query: 1 MQLLEERILTDGUVLGENILKVDSPFLTHQVDFELMQEIGKVPADKYKEAGITKVVTIEAS 60
 M+LLEERIL DG++LGENILKVD+PLTHQVD+ LM+ IGKVPFA KY EAGITKVVTIEAS
 Sbjct: 1 MQLLEERILTDGILLGENILKVDNPLTHQVDFELMQEIGKVPADKYKEAGITKVVTIEAS 60

-2410-

Query: 61 GIAPAVYAAQALGVPMIFAKKAKNITMTBGLITAEVVSPTKQVTSQSIVSRFLSNDTV 120
 GIAPAVYAA+ A+ VPMIFARK KNITMTBGLITAEVVSPTKQVTS VSI +FLS +D V
 Sbjct: 61 GIAPAVYAAEADVMIFAKKHNITMTBGLITAEVVSPTKQVTSVSLAGKFLSDEKV 120

Query: 121 LIIDDFLANGQAAGKGLSIIGQAQAGKAVAGIVIEKSFQDGRDLLEKTGVPTSLARIKA 180
 LIIDDFLANGQAAGKGLSIIGQAQAG+V G+GIVIEKSFQDGR L+E G+ VTSLARIK
 Sbjct: 121 LIIDDFLANGQAAGKGLSIIGQAQAGVVGVSIVIEKSFQDGRRLIEMGRVTSLARIKN 180

Query: 181 FENG RVV PARADA 193
 FENG + F EADA
 Sbjct: 181 FENG NLFLEADA 193

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2141

A DNA sequence (GBSx2257) was identified in *S. agalactiae* <SEQ ID 6615> which encodes the amino acid sequence <SEQ ID 6616>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2546 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB15203 GB: Z99120 similar to GMP reductase [Bacillus subtilis]
 Identities = 243/321 (75%), Positives = 286/321 (88%), Gaps = 2/321 (0%)

Query: 7 VFYEDIQLIPNKCISSRSQADTSVKLGNYTFKLPV+PANMQTIDEVETLACEGYF 66
 VFYEDIQLIP KCI++SR+ DTSV+LG +TFKLPV+PANMQTIDE++A +LA GYF
 Sbjct: 4 VFYEDIQLIPAKCIYNSRSECDTSVRLGGHTFKLPVVPANMQTIDEKLAIslaENGYF 63

Query: 67 YIMHRFNEEERKPFIKRMDKGLIASISVGKIDYEYFVTSLEED--APEFTITDIAHGH 124
 Y+MHRF E R FIK N+ +GL +SISVGKID EY+FY L E+ FE++TIDIAHGH
 Sbjct: 64 YIMHRFPEETRDIKIDKARGLFSSISVGKIDEEYEFVQLAEENLTPEYVTDIAHGH 123

Query: 125 SNSVIMIQHIKQELPETFVFIAGNVGTFPAVRELENAGADATKVGIGPGKVCITKIKTF 184
 SN+VIMIQH+K+ LP++FVIAGNVGTFPAVRELENAGADATKVGIGPGKVCITK+KIKTF
 Sbjct: 124 SNAVIMIQHLKKHLPDSFVIAGNVGTFPAVRELENAGADATKVGIGPGKVCITKIKTF 183

Query: 185 GTGGVQLAALRWCSKAARKPIIADGGIRTHSDIAKSIRPGASNMWIGSLFAGHLESFGKL 244
 GTGGVQLAALRWCKAA KPIIADGGIRTHSDIAKSIRPGA+NMWIGSLFAGH ESFG+
 Sbjct: 184 GTGGVQLAALRWCAKAAKPIIADGGIRTHSDIAKSIRPGATNMWIGSLFAGHLESFGQT 243

Query: 245 VEVEGQCPKEYYGSASEYQNGHHNVGEGKILLPVKGRLEDTLTFMQDQLQSSISYAGGK 304
 +E +G+ +KEY+GSASE+ RGH KNVEGKK+ + KQ ++DTL EK+QDLQSSISYAGG
 Sbjct: 244 IEDKDKLYKEYYGSASEFPAGCKKNVGGKMMHVAHGGSIKDTLLEMQDQLQSSISYAGGT 303

Query: 305 ELDLSRHVDYVIVKNSIWNKD 325
 +L++R+VDYVIVKNSI+NGD
 Sbjct: 304 KLNALRNVDYVIVKNSIFNKD 324

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6617> which encodes the amino acid sequence <SEQ ID 6618>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2411-

bacterial cytoplasm --- Certainty=0.2405 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 297/327 (90%), Positives = 311/327 (94%)

Query: 1 MFNDIPVFDYEDIQLIPNKCIISSRSQAUTSVKLGNYTFKLPVIPANMQTIIDBEVATL 60
 10 MFNDIPVFDYEDIQLIPNKCII+SRQAUTSV LG Y FKLVPVIPANMQTIIDR +AE L
 Sbjct: 8 MFNDIPVFDYEDIQLIPNKCII+SRQAUTSVTLGKYQKLPVIPANMQTIIDRTIAPQL 67

Query: 61 ACEGYFYIMHRFPNEERKPFKIRMHDKGLIASISVGVDYEDFVTSLKEDAPEFITIDI 120
 A EGYFYIMHRF+E+ RKPFIKRMH++GLIASISVGVK EY+FVTSLKEDAPEFITIDI
 Sbjct: 68 AKEGYFYIMHRFDSDSRKPFKIRMHEQGLIASISVGVKACEYEFVTSLKEDAPEFITIDI 127

15 Query: 121 AHGHSNSVIMIQIKQLPETPVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 180
 AHGH+NSVI+MI+HIK ELPETPVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV
 Sbjct: 128 AHGHSNSVIDMIKHITQLPETPVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 187

20 Query: 181 KTGFGTGGWLAALRWCSKAARKPIIADGGIRKTHGDIANSIRFGASVMVIGSLFAGHLES 240
 KTGFGTGGWLAALRW+C- KAARKPIIADGGIRKTHGDIANSIRFGASVMVIGSLFAGH ES
 Sbjct: 188 KTGFGTGGWLAALRWCAKARKPIIADGGIRKTHGDIANSIRFGASVMVIGSLFAGHLES 247

25 Query: 241 PGKLEVEGQQFKEYYSASEYQGEHKNVGEKILLPVGRLEDLTLEMQQDLQSSISY 300
 PGK VEV+G+ FKEYYSASEYQGEHKNVGEKILLP KG L DLTLEMQQDLQSSISY
 25 Sbjct: 248 PGKTVVDGDTFKEYYSASEYQGEHKNVGEKILLPTEGHLSDLTLEMQQDLQSSISY 307

Query: 301 AGGKELDSLHRHVDYIVIVKNSIWNQDSI 327
 AGGK+LDSLHRHVDYIVIVKNSIWNQDSI
 30 Sbjct: 308 AGGKELDSLHRHVDYIVIVKNSIWNQDSI 334

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2142

35 A DNA sequence (GBSx2258) was identified in *S.agalactiae* <SEQ ID 6619> which encodes the amino acid sequence <SEQ ID 6620>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have an uncleavable N-term signal seq

40	INTEGRAL	Likelihood = -16.98	Transmembrane	421 - 437 (413 - 443)
	INTEGRAL	Likelihood = -8.81	Transmembrane	166 - 182 (159 - 186)
	INTEGRAL	Likelihood = -8.55	Transmembrane	220 - 236 (208 - 238)
	INTEGRAL	Likelihood = -6.69	Transmembrane	322 - 338 (319 - 353)
	INTEGRAL	Likelihood = -5.26	Transmembrane	199 - 215 (196 - 218)
45	INTEGRAL	Likelihood = -4.35	Transmembrane	343 - 359 (342 - 361)
	INTEGRAL	Likelihood = -4.09	Transmembrane	291 - 307 (287 - 308)
	INTEGRAL	Likelihood = -3.66	Transmembrane	8 - 24 (8 - 27)
	INTEGRAL	Likelihood = -3.66	Transmembrane	133 - 149 (133 - 151)
	INTEGRAL	Likelihood = -3.19	Transmembrane	254 - 270 (253 - 278)
50	INTEGRAL	Likelihood = -2.50	Transmembrane	53 - 69 (53 - 72)
	INTEGRAL	Likelihood = -1.81	Transmembrane	77 - 93 (76 - 95)
	INTEGRAL	Likelihood = -1.33	Transmembrane	109 - 125 (109 - 125)

----- Final Results -----

55 bacterial membrane --- Certainty=0.7793 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:CB61253 GB:AJ250422 ORFC [Oenococcus oeni]
 Identities = 157/447 (35%), Positives = 252/447 (56%), Gaps = 13/447 (2%)

-2412-

Query: 11 AIITTAIGFSGILIEISRMNVTFLLMKRGVNPVAVIQWVITGNLLAVAVTVPLSAFMK 70
 AI+ A L F G+LIETSRMNVTF LM++F ++ +QN+IT LL VA T+ ++AF+ K
 Sbjct: 15 AILGLAGLAPCOVLISRMNVTFITLMQCFSLISLNKVKQMLTAYLLVAATISDAFIER 74

5 Query: 71 NLSEKQIFTLANVFLVSGVLIDSAPNLAILLVGRVLQGVGTGLAPLLFHITAIQIME 130
 ++IF A +LF+ GV+ + AEN ILL+GR++Q + TGLA+PLL I+ QIP +
 Sbjct: 75 RPIPKKIFFWAGLLFIIGVICSAIAFNFLILLIGRLIQALSTGLAIPLLITEIMQKIPK 134

10 Query: 131 RRLMGVAMVMTLLAPAVGPTYGGVISGMLGWMKIMPLLAFLIISTFGLASIPKRQV 190
 ++G M+ + L P++GPTYGGVI+ L W++IF + PI +I+ IGL+ I ++
 Sbjct: 135 KQSYMELVWILLWQPSLQPTGYGVITQDLSWRLLFWFVLPIGLIAMLIGLSFIBQKSS 194

Query: 191 RINDKINFPFISLIGLITALLAIEKMSIP-----YLLVAIVSFVIFYL--NRK 239
 + + FISL+ L ++ +A+ I+ +LL+A++ ++F L N +
 15 Sbjct: 195 PSKIPFANKQFISLILALLSITVAVNNAGIYGMTSIFKFGFLIAVILLIVFIKLNSR 254

Query: 240 LEPLNINVPKDKFSILLYGVLAFCQMIPALSLFLPNLLQVILHCTSTKAGLFMFGAIA 299
 ++++++FK +F L Q I L+L+FLGPN QI+L + +G+ + G++
 20 Sbjct: 255 QALISISIPKMEFVCPMLLIYFLIQSLTFLPNYALQLIKKGVMSIGIMLQCSLI 314

Query: 300 VVFLSPFAGYLLDKIGAFKPIMIGISLSLIGLIGTAIFIPAKSVVLLAFDILLKIMG 359
 L P G +LD P++IG + I IF SV ++ A ++ IG
 25 Sbjct: 315 SAILQPLTGRHLDSPVKIPLVIGAFPLITSTISFTIPQRLSVFLAALYVTIMGFSF 374

Query: 360 GASRMVMTALTILKLPAGSADGNSILMTLQQFAGAFATVAVSQIFITGQVAIPKNGAIGS 419
 +N +T AL KL +DGN++ NTLQQ+AG+ T+VAS + G K GS
 30 Sbjct: 375 VFNSLTYALQKLPKLKILSDGNVFTLQQYAGSLGTSVASALLANGIGTDGKQSYTGS 434

Query: 420 Q--FAVLVIVVIVLAIVGLTYLKRK 444
 + P + F+ +++ ++ +K K
 30 Sbjct: 435 RHIFILNFISCAITVILIFSIQRKKNK 461

There is also homology to SEQ ID 46.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2143

A DNA sequence (GBSx2259) was identified in *S.galactiae* <SEQ ID 6621> which encodes the amino acid sequence <SEQ ID 6622>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2151(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6595> which encodes the amino acid sequence <SEQ ID 6596>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2413-

Identities = 74/214 (34%), Positives = 112/214 (51%), Gaps = 5/214 (2%)

Query: 13 MESENPFITLACTYFNYLPSIQIT---DISTLNHADPGSPAFPHDITSIPIHVAIDSNY 69
 N+ E F L +F+LF +I+T +I + + F G F+H+ +P L ++
 5 Sbjct: 15 NQLBETTFIRELSHHFSLFEVITLTSKANTQSNQLSTFGGIFSPHEDIDLFTLYFKTSQ 74

Query: 70 LAISQINSKIRANDIKTFSSELKMTIEFHYMLNDFLPHLPYRFLINKGQGTYSNEIKP 129
 ++ + LS+ +T F+ + +LP + RL + +G I NH
 10 Sbjct: 75 HQGSLVTRSVFDQRTAVLSLSQYLTGFTYKQFGHFLQYLPLQARLSDAINGITLVNNAF 134

Query: 130 EDPFDIYPESEYFIDKQWQNSLIEKKAKELHLLPSASQVILVCSYKRLNDGQVACY 189
 F F + I+ W+ L LLS S D+I +Q Y+ L+N QQLGQ
 10 Sbjct: 135 NGSF--LFTTDKETEHWLAELRLSDNPCKTFLPSSGLDHYMCHYQALKNPQSLGV 192

Query: 190 IEHVNRIKPLEGLYKESGQAIVGMSDVTSGASI 223
 ++ V +IKPLL YL+B+QQAIVGMSDVTSG SI
 15 Sbjct: 193 LDTVQDIKPLANQLEETGQAIVGMSDVTSGPSI 226

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2144

A DNA sequence (GBSx2260) was identified in *S. agalactiae* <SEQ ID 6623> which encodes the amino acid sequence <SEQ ID 6624>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.10	Transmembrane	431 - 447 (423 - 452)
INTEGRAL	Likelihood = -8.92	Transmembrane	149 - 165 (147 - 174)
INTEGRAL	Likelihood = -8.86	Transmembrane	404 - 420 (402 - 428)
INTEGRAL	Likelihood = -7.91	Transmembrane	299 - 315 (293 - 316)
INTEGRAL	Likelihood = -6.42	Transmembrane	380 - 396 (374 - 398)
INTEGRAL	Likelihood = -5.31	Transmembrane	350 - 366 (347 - 367)
INTEGRAL	Likelihood = -4.57	Transmembrane	56 - 72 (54 - 74)
INTEGRAL	Likelihood = -3.24	Transmembrane	172 - 188 (171 - 196)
INTEGRAL	Likelihood = -1.33	Transmembrane	224 - 240 (224 - 240)
INTEGRAL	Likelihood = -0.59	Transmembrane	101 - 117 (101 - 117)

----- Final Results -----
 bacterial membrane --- Certainty=0.5840 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF84709 GB:AE004010 potassium uptake protein [Xylella fastidiosa]
 45 Identities = 201/570 (35%), Positives = 319/570 (55%), Gaps = 34/570 (5%)

Query: 1 NASMQHVNHSSFDKAKAGFII--ALGIVYGDIGTSPLTYMQSLVENQGGISSVTSFIL 58
 N+ H + ++ G II A+G+V+GDIGTSPLT++ G++ ++ ++L
 50 Sbjct: 1 NSTSSHSQDCTAVPSNKSITLLSALGVVFGDIDIGTSPLTYLKEAFSPNYGLTPRHDT-VL 59

Query: 59 GSISLIWTLTLITTIKVVLWALKADNHHS33IFSLYTLVRKMTFW-----LIVFAVI 111
 G +SLI W + L+ TKIV V ++ DN EGGI +L L ++ P+ + + +
 50 Sbjct: 60 GILSLIFAMHMLVVTIKYVAVIMRVNDGEGGIMALTALQRTMPFGSRSTIYVGLHGF 119

Query: 112 GGATILSGALTEAVTVTSVAHGKVVVPSLQHPQNSVIFATLFIILLFAIQRFGTG 171
 G + DG +TER++V SAVGCL+V F V+ TL +L+LLF QRFPT
 55 Sbjct: 120 GTSLEFGDGVITPAISVLSRAVEGLEVAEPHMKAF-----VVPITLAVLILLFLCQRFQTE 174

Query: 172 VICKLFGPTMPINFAFLGSLINSFAHPEVFKAINPYTGLKLFSPNNHKGIFILGSIF 231
 +CK FGPT ++WF +G+ G+ N PEV AINE +GL F +F+LG++
 60 Sbjct: 175 RVCKTFGPITLLWFALGVGVYNTIAQAPEVLHAINPSGLH-FFLEHGHSMFVLGVV 233

-2414-

- Query: 232 LATTGAEALYSDLGHVGRGNHVSFWFKVAII-LSYCGQQAAILANVAGNELNPPFPAS 290
 LA TG EALY+D+GH G I +W +V + + L+Y CQGA +L+N A NPP+ S
 Sbjct: 234 LAVTGGEALYADMSHFCAIKAIHANNVVLPMALANTLQAGVALSNPTAIG--NFFTQS 291
- 5 Query: 291 IPGQTMHVILATLAALISQALISGSPFLVSEAMRLKIPFQFSTYTGLEN-IGQTYIP 349
 IP ++ LAT AA-IAASQAL+GS++L S+M+L P+ + + IGQ Y+P
 Sbjct: 292 IPDWGLYPMIALATRAAVIASQALITGSYSLSQAMQLGYPRMNVHSTQSSTIGQIVP 351
- 10 Query: 350 VINWFLFAITTSIVLLFKTSRAMEAAYGLAITIMLMTILLSPFL-IQNGVKRGLVLLM 408
 +NW L + V+ F S M +AIG+A+T IM++TT+L+ + V R ++ +M
 Sbjct: 352 TVNWTLTLVILITVIGFGDSTMSASAGVAVTGMHITTVLMIIYARANPRVRLMARM 411
- 15 Query: 409 MIFPGILEGIFFLASAVKFMHGGYVVVILAVAILFIMTIWYKGSKIVSRVVL--LLKD 466
 I F ++G FF A+ +KFM G + ++ V I M W +G K++ + + ++L +
 Sbjct: 412 AIVFIADVGAFFYANIIKFMGDAWFFLLGVVIFIPMKTWLGRKLLHEMRKKGINLON 471
- 20 Query: 467 YIGQLDKLRHHRYPYIHTNVVYLITNMEEDMDKSMYSILDKRPFKGAQVYWFVNKVT 526
 ++ L L + P V+LT + ++ +W++ + + + F+ +K
 Sbjct: 472 FPLGL-MLAPPVKV---GVAVLT--ADSTVVPHALMNLKHNKVLHERNV-FLTVKTL 524
- 25 Query: 527 DEPYTA---EYKVDMMGTDFIVKVELYLGF 553
 PY A K++ + F + GF
 Sbjct: 525 KIPTAANSERLKIEPISNGF-YRVHIRFGF 553
- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6625> which encodes the amino acid sequence <SEQ ID 6626>. Analysis of this protein sequence reveals the following:
- Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq
- 30 INTEGRAL Likelihood = -11.78 Transmembrane 428 - 444 (421 - 453)
 INTEGRAL Likelihood = -8.70 Transmembrane 146 - 162 (144 - 171)
 INTEGRAL Likelihood = -7.64 Transmembrane 404 - 420 (398 - 426)
 INTEGRAL Likelihood = -4.88 Transmembrane 296 - 312 (294 - 315)
 INTEGRAL Likelihood = -4.57 Transmembrane 53 - 69 (51 - 71)
 INTEGRAL Likelihood = -3.93 Transmembrane 347 - 363 (343 - 363)
- 35 INTEGRAL Likelihood = -2.50 Transmembrane 372 - 388 (371 - 388)
 INTEGRAL Likelihood = -1.33 Transmembrane 169 - 185 (169 - 185)
 INTEGRAL Likelihood = -1.33 Transmembrane 221 - 237 (221 - 237)
- Final Results -----
- 40 bacterial membrane --- Certainty=0.5713 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- The protein has homology with the following sequences in the databases:
- 45 >GP:AAF84709 GB:AB004010 potassium uptake protein [Xylella
 fastidiosus]
 Identities = 177/467 (37%), Positives = 270/467 (56%), Gaps = 20/467 (4%)
- 50 Query: 7 TAPFKASKAGPII-ALGIVYGDIGTSPLYTIQSLVENQGVQKVSFPIQSILSIWTL 65
 TA S I+ A+G+V+QDIGTSPLYT++ G+ ++ +G +SLI W +
 Sbjct: 11 TAVPNSNGTIIILSAIGCVVPGDIGTSPLYTLKAFSPNYGLTPNHDT-VGLIISLIFAM 69
- 55 Query: 66 LTITTKYVILALKADNHHGGGIFSLPTLVKRMSPW-----LIIPMIGCATLLSDCA 118
 L+ TTKYV ++ + DN EGGI +L L ++ P+ + I + G + DG
 Sbjct: 70 MLVVTKYVAVIMRVNDNDEGGGIMALTALQTQTMPPGSRSIYVIGIIPGTSFLPGDGV 129
- 60 Query: 119 LEPAVTVTSALTEGLKAVPGLSHIYQNGINVIITLVILVLFGIQRFGPGFKIGKIPFM 178
 +TFA+V SA+EGLA + V+ TL +LI+LF ORRT +GK RQ+
 Sbjct: 130 ITPAISLVSAVEGLEVAEPHMAF-----VVPITLAVLILFLQCRPGTNRVGTGTPIT 184
- 65 Query: 179 PIWFSFLGVSFPNTLGHLEIKAIPIYPYALHLPSPENHRGIFILGSLFIATTAAREALY 238
 +WF +GV G +N E+ AIMP + LH F +F+IG+ LA TG EALY
 Sbjct: 185 LLWFIATGVGVYNIQAPEVLHAINPSCGLH--PFLHGWHSMPVLCAVNAVLTQGEALY 243
- Query: 239 SDLGHVGRGNITVSWPFKVM-CIVLSYCGQAAILANKHSGITLNPFPASVSPQLRVLYLV 297

-2415-

+D+GH G I +W +V + + L+Y GQ A +L+N + NPF+ S+P ++
 Sbjct: 244 ADMGHFGAKIRHAWRYVLIHMLALNLYLQGGALVLSNPTA--IGNPFYQSIPDWGLYPMI 301

5 Query: 298 SLATLAATIASQALISGSFTLVSEAMRLKIPFLPRVITYPG-ANLQQLYIPVINWILFAVT 356
 +LAT AA+IASQALI+GS+L S+AM+L P V + + +GQ+Y+P +NW L +
 Sbjct: 302 ALATAAAVIASQALITGSYLSLSSQAMQLGYIPRMNVHTSQTIGQYVYPTVNNWILLTLV 361

10 Query: 357 SCTVLAFTSAHMEAAAYGLAITITMLMTILLKYLLIKKGRTPILAHLMVAF-PAIVEFI 415
 TV+ F S M +AYG+A+T TM++TT+L+ Y P L +MA P+
 Sbjct: 362 ILTVIGFGDSTSMASAGYAVIGTMMITTVMITITVARANPRVPRMLMMAVPIAVDGA 421

15 Query: 416 FFLASAIFKPMHGGYAVVILALAIIVPMFIMHAGTRIVFKYVKSINLN 462
 FF A+ IKPM G + ++L + I M W G +++ + ++ +N
 Sbjct: 422 FFYANIIFKPMGAWFFLLGLGVVITTFMRITWLGKRLKHEEMRKGIN 468

An alignment of the GAS and GBS proteins is shown below.
 Identities = 485/651 (74%), Positives = 575/651 (87%)

20 Query: 10 SSFDKASKAGPIIALGIVYGDIGTSFLYTMQSLVENQGGISVTSFILGSI SLIITWLT 69
 ++FDKASKAGPIIALGIVYGDIGTSFLYT+QSLVENQGG++ V+BSFILGSI SLIITWLT
 Sbjct: 7 TAFDKASKAGPIIALGIVYGDIGTSFLYTIQSLVENQGGVNVQV+BSFILGSI SLIITWLT 66

25 Query: 70 LITTIKYVLVALKADNRHHEGGIFSLYTLVRKMTFNLIVFAVIGGATLSDGALTPAVT 129
 LITTIKYVL+ALKADNRHHEGGIFSL+TLVRKM+PMLI+PA+IGGATLSDGALTPAVT
 Sbjct: 67 LITTIKYVLIALKADNRHHEGGIFSLFTLVRKMSFNLIIPAMCGGATLSDGALTPAVT 126

30 Query: 130 SAVEGLKVPSLQHFQNGENVIFATFLIMILLFAIQRFQTVIGIKLPGPINFIFWAF 189
 SA+EGLK VP L HI+QNG+VNI TL IL++LF IQRFQIG IKG+PGP+MFIWF+FLG
 Sbjct: 127 SAIEGLKAVPGLSHYQNGTNNVITTLVILVILFGIQRFQIGFICKPGPVMFVWFSFLG 186

35 Query: 190 ISGLLNSFAHPEVFKAINFYQJLKLSPENHKGIFILGSIPLATGAALYSDLGHVGR 249
 +SG N+ H+EKAINFY L LSPENH+GIFILGSIPLATGAALYSDLGHVGR
 Sbjct: 187 VSGFPNTLGHLEIFKAINFYALHLLSPENHKGIFILGSIPLATGAALYSDLGHVGR 246

40 Query: 250 GNIHVSWPFVKVAILSYCCQAWILANKHMGAGNELNPFPSISQPTMHWILATLAAII 309
 GNI+VSWPFVK+ I+LSYCCQ AWILANK++G ELNPFPS+PSQ +++V LATLAAII
 Sbjct: 247 GNIYVSWPFVOMCIVLSYCCQAAWILANKHSGIELNPFPSVPSQLRYVLVSLATLAAII 306

45 Query: 310 ASQALISGSFTLVSEAMRLKIPFPRSTYPODNIQCTYIPVINWFLPAITTSIVLLFKTS 369
 ASQALISGSFTLVSEAMRLKIPF PR TYPG N+GQ YIPVINW LPA+T+ VL P+TS
 Sbjct: 307 ASQALISGSFTLVSEAMRLKIPFLPRVITYPGANLQQLYIPVINWILFAVTSCTVLAFTS 366

50 Query: 370 AHMEAAAYGLAITITMLMTITLLSFFLIQKQKVRGLVLLMMIFPGILGIPFLASAVKPMH 429
 AHMEAAAYGLAITITMLMTITLL ++LI+KG + L L+M FF ++E IFFLASA+KPMH
 Sbjct: 367 AHMEAAAYGLAITITMLMTITLLKYLLIKKGRTPILAHLMVAFPAIVEPIFLASAIKPMH 426

55 Query: 430 GGYVVVIVAVAIIFIMTINWYKSKIVSRVYKLLDLKDYIQGLDKLRDHRYPIYHNVVY 489
 GGY VVI+A+AI+P+M IN+ G++IV +YVK L+L DY Q+ +IR D + +Y TNVVY
 Sbjct: 427 GGYAVVILALAIIVPMFIMHAGTRIVFKYVKSINLNDYKQKIQKLRDDVCDLYQTNVVY 486

60 Query: 490 LTRNMEEDMDKSDMSYILDKRKPQAVVYVFNKIVTDEPYTAAYKVDNMGTDPIVVKEL 549
 L+NRN++ MID+SI+YILDKRKP+AQVYVFN++VIDEYTTA+YKVDNMGTD++V+V L
 Sbjct: 487 LSNRWQDHMDRSILYSILDKRKPQAVVYVFNQVTDDEPYTAAYKVDNMGTDYVYVNL 546

65 Query: 550 YLGFQMRQPTVSRYLRTIVIEILRSGRLPKQKGYTSVRPDSNVGQDPRFVILDERPSSQNL 609
 YLGF+QNT QYI RLRTIV++L+BSGRLPKQ +Y++ P +VGDFRP++++ER S+++ L
 Sbjct: 547 YLGFQMRQPTVSRYLRTIVQDILMSGRLPKQKQYETITPGRDVGDPRFVILDERVSNARQL 606

60 Query: 610 KPFGRFVLMKSSIKHWATIPRWFLQFSEVTEVVEVPLIIPANKLPIKE 660
 ERP+M K+SIKH TA+P+RWFLQ+SEVT EVVPLI + LPKE
 Sbjct: 607 SNFERFIMCTKASIKHWATAPRWVFLQYSEVTEVVEVPLIILSDVLKLPKE 657

A related GBS gene <SEQ ID 8983> and protein <SEQ ID 8984> were also identified. Analysis of this protein sequence reveals the following:

65 Lipop: Possible site: -1 Crend: 8

-2417-

```

          360      370      380      390      400      410      420
1590      1620      1650      1680      1710      1740      1770      1800
5  LASAVKFMHGGYVVVITAVATITPMTITWYKSGKIVRYVKLLDLKDYIGQLDKLRHDKHYTYHTNVVYLTNRMEEDMID
   |:||| |:| |:| |:| |:| |:| |:| |:|
YANITIKPMIDGAWFLLLGVVITPMTITWYKSGKIVRYVKLLDLKDYIGQLDKLRHDKHYTYHTNVVYLTNRMEEDMID
          440      450      460      470      480      490      500

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2145

A DNA sequence (GBSx2261) was identified in *S. agalactiae* <SEQ ID 6627> which encodes the amino acid sequence <SEQ ID 6628>. This protein is predicted to be serine dehydrogenase. Analysis of this protein sequence reveals the following:

```

15  Possible site: 26
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.3261 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

25  >GF:ARD07424 GB:AE000552 short chain alcohol dehydrogenase
    [Helicobacter pylori 26695]
    Identities = 10/31 (58%), Positives = 25/31 (80%)

    Query: 3  NVASQPEHINIRIEMFVSQTYGQPFVYED 33
           W+  QP HANIRIK+NEVSQ+YGPQFV ED

30  Sbjct: 219 WIYQPEHINIRIEMFVSQTYGQPFVYED 249

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6629> which encodes the amino acid sequence <SEQ ID 6630>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 21
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.1021 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 24/33 (72%), Positives = 29/33 (87%)

45  Query: 1  MSWVASQPEHINIRIEMFVSQTYGQPFVYED 33
           +SWV  QP HANIRIK+NEVSQ+YGPQFV ED

      Sbjct: 20 VSWVHQPFHNVANIRIEMFVSQTYGQPFVYED 52

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2146

A DNA sequence (GBSx2262) was identified in *S. agalactiae* <SEQ ID 6631> which encodes the amino acid sequence <SEQ ID 6632>. Analysis of this protein sequence reveals the following:

-2418-

Possible site: 21
 >>> May be a lipoprotein

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9337> which encodes amino acid sequence <SEQ ID 9338>
 10 was also identified. A further related GBS nucleic acid sequence <SEQ ID 10781> which encodes amino acid sequence <SEQ ID 10782> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10951> which encodes amino acid sequence <SEQ ID 10952> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CRA32349 GB:X14130 ORF (AA 1 to 299) [Lactococcus lactis subsp. cremoris]
 Identities = 72/215 (33%), Positives = 110/215 (50%), Gaps = 8/215 (3%)
 Query: 4 RSKLAAGFLTMSVATLAACSGKTSNGIN--VVMKGDITITVSDFIDOVKISKAAQQSM L 61
 + K+ L + L SG SN T+ V T G +T S FY ++K S + +
 20 Sbjet: 2 KKIQRRLKLLASTATALLLLGGCGSNCTDQTATYSGGKVTSESSFYKELKQSPPTKIMLA 61
 Query: 62 TLILSRVFDTCYGDVKDDKKVSEAYNKATKGYGNSFSSALSGAGLTPGYKQIRPTIMLV 121
 +++ R + YG VS K V++AY+ + YG +F + LSQ G + +K+ +RT L
 25 Sbjet: 62 NLLIYRALNHAYGKSVSTKTVNDAYDSYKQYGENFDALFSGNGFSRSFKESLRTNLFSL 121
 Query: 122 EYAVKEAAKELTANYKYAYNYTPETSVQVIKLDAEDKAKSVLKVDKAGDAFAKIAK 181
 E A+K+ K+++E+ K A+K Y P+ +VQ I ED AK V+ D+ A G DPA +AK
 Sbjet: 122 EVALKKL--KKVBSQQLAAWKTYQPKVTQHLLTSDDETAKQVISDLAA-GKDFAMIAK 178
 30 Query: 182 E---KTTATDKKVEYKFDAGITLPEKVMSSAFKL 213
 T D + F+ TL AA+KL
 Sbjet: 179 TDSIDATKDKNGKISFELNNKTLDAATFKDAAYK 213

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6633> which encodes the amino acid
 35 sequence <SEQ ID 6634>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> May be a lipoprotein

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the databases:

>GP:AAA25247 GB:M83946 maturation protein [Lactobacillus paracasei]
 Identities = 88/294 (29%), Positives = 146/294 (48%), Gaps = 14/294 (4%)
 Query: 7 LIASVVTLASVWALAAQCSQNDNIRKVISNKGDTISVSDFYNETKNTVEVSQKAMNLIVSR 66
 L+AS T +++ L+ QGS + KV + G ++ S+FY E K + ++ +N+I R
 50 Sbjet: 10 LLASTAT--ALLLLSGCSQNDADQKVATYSGGKVTSESSFYKELKQSPPTKIMLANMLYR 67
 Query: 67 VFEEAQYGDVKVSKKEVEKAYIKTARQYGASFSALLQSLITPETFKRISSKLVETAYKE 126
 YG VS K V AY +CYG +F A L+Q+ + +FK +R++ L E A+K+
 55 Sbjet: 68 ALNHAYGKSVSTKTVNDAYDSYKQYGENFDALFSGNGFSRSFKESLRTNLFSLVALAK 127
 Query: 127 AAKKELITQEYKAYEYSTPTMAVEMITLONEETAKSVLEKLKARGADPTAIAKE---KT 183
 K+++ + K +++Y P + V+ I +E+YAK V+ +L A G DF +AK T
 Sbjet: 128 L--KKVBSQQLAAWKTYQPKVTQHLLTSDDETAKQVISDL--AAGKDFATLAKTSDIT 184
 60 Query: 184 TTFEKKYTKYKFDGATNVPTDVKRAASSINEGGISDVISVLDPTISYKQKPKVIVKTKKAE 243

-2419-

T + F+S + AA L Q + P + ++K+
 Sbjct: 185 ATKNGGKISFESNNKTLDRTPKDAAYKLNGDITQT-----PVKINGYEIVIMNH-P 238
 Query: 244 KKSDDQIEYKPLKAILIAEKSDWNPQRKIVIANALDKANVKIKDKAPANILAQY 297
 K + KK L A + A+ S+D + +VI+ L +V IKDK A+ L Y
 Sbjct: 239 AKGTPSSKALTASVYAKWSDSSINQRVISQVLKKNQVHTIKDKDLADALDSY 292

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/213 (58%), Positives = 168/213 (78%), Gaps = 1/213 (0%)
 Query: 1 MTRSRKLAAGPLTMSVAITLAACSGKTSNOTNVVMKGGDTITVSDFYDQVKTSKAAQSN 60
 MK +KL A +TL SV LAAC T++ T V++MKGGDTI+VSDFY++ K ++ +Q+K+
 Sbjct: 1 MKNSNKLIAASVITLASVMAALAAQS-TMDNTKVISMKGGDTISVDFYNETKNTVEYSQKAM 59
 Query: 61 LTLILSRVFDIQYGDVKDKVSEAYNKTKAGYGNFSFSAASQGLTPEGYKQIQTITML 120
 L L+SRVF+ QYGDKVS K+V +AY+KTA+ YG SPS+AI+Q+ LTPR +K+QIR++ L
 Sbjct: 60 LNLVLSRVFPAQYGDVKSKVEBKAYHKTABQYGASPSAALAAQSSLTPETFKRQIRSEKL 119
 Query: 121 VEYAVKEAAKKELTEANYKEAYKNYTPETSVQVVKLDAEDKAKSVLKDVKADGADFARIA 180
 VEYAVKEAAKKELT YK+AY+YTP +V+I LD E+ AKSVL++KA+GADF IA
 Sbjct: 120 VEYAVKEAAKKELTTQETKKAYSYPTPMABMITLONEETAKSVLEELKABGADFTATA 179
 Query: 181 KEKTTATDKKVEYKFDASGATTLFKEVMSAAPKL 213
 KEKIT +KKY YKFS T +P +V AA L
 Sbjct: 180 KEKITTEKKVTKYKFDGATNVPTDVKAASSL 212

SEQ ID 10782 (GBS657) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 8-10; MW 62.8kDa) and in Figure 187 (lane 3; MW 63kDa). Purified GBS657-GST is shown in Figure 245, lanes 2 & 3.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2147

A DNA sequence (GBSx2263) was identified in *S. agalactiae* <SEQ ID 6635> which encodes the amino acid sequence <SEQ ID 6636>. This protein is predicted to be methyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2576 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]
 Identities = 132/227 (58%), Positives = 169/227 (74%)
 Query: 1 MVQSYKSNANHMRRPVVKEIVCYMRQHOKQNNCCLAELFAKQBNIPPIPHETATYP 60
 MV+Y +N M RPVVK E+V++MR Q Q Q LAE+ FAK+ NIP+IPIET YF
 Sbjct: 1 MVETKYKTSNPNMRRPVVKAELVEMRRSSQTVIGELAEVLNFAKKNIPVPIPHETVLYF 60
 Query: 61 RFLMOTLQPKHILEICTAIGFSAALMAENAPAKITTIDRNBEMIALAKENFAKYDNNQ 120
 + L+ L+PK ILRIGTAIGFSA+MA+ PEA+I TIDRN EMI LAK+N AKYD+ NQ
 Sbjct: 61 QMLLSLKPFRILRIGTAIGFSAVMAQEVPEABIVTIDRNPHEMILAKKNIAKYDNNQ 120
 Query: 121 ITLLEGDADVDTLQLOKSYDFVMSAKSKYIVFLPQVLKHLDDUGGVVLLDDIFQGGDIA 180
 I L EGDA DVLQ L +D VFMSAKSKY+ FLP+ L+ L G+++DD+FP G+I
 Sbjct: 121 IQLKSGDADVQLQLKGPDPFLVMSAKSKYVFLPKSLLELSNGLILMDDVFPQGEIL 180

-2420-

Query: 181 KPIDEVRRGQRTIYRGLQLRLEFSTLQHPDLTATLVPLDGLLMIRKN 227
 PI EV+R QR + RSL+LFD +P +++PLDGLLMI+K+
 Sbjct: 181 LPIMEVRRNRQALRGLRGLRLEFDFVDPNPKYMTSVLPLDGLLMIKKH 227

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6637> which encodes the amino acid sequence <SEQ ID 6638>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 10 INTPGRAL Likelihood = -1.38 Transmembrane 153 - 169 (152 - 170)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]
 Identities = 134/227 (59%), Positives = 169/227 (74%)
 20 Query: 1 MVKSYSTANHNMRPVPVKELVHYMRTQKQITGFLAELEQFARQENIPIIQPEVVAIF 60
 MV+Y T+N M RPVK ELV +MR+ Q Q TO LAE+ FA++ NIP+I E V YF
 Sbjct: 1 MVETKYSTSNPMNRPVPVKAELVEMRSSQTQVTSELAEVLNFAKNNIPIVPIHETVLYIF 60
 25 Query: 61 RFLQLSQPKHILEIGTAIGFSALLMAENAPDATTITVTDNRNEMIDPAKANFAKYDSRQ 120
 + LL L+PK ILEIGTAIGFSAL+HA+ P+A IVTIDRN EMI+ AK N AKYD R Q
 Sbjct: 61 QMLLSLLKPKHILEIGTAIGFSALVQAQVPEAEIVTIDNRNEMIELAKKNLAKYDHRNQ 120
 30 Query: 121 IRLLEGDAADILSTLEGNFDFVPMDSAKSKYIVFLPEILRLLLKVGGVVLLDVPQSGDIT 180
 I+L EGDAD+L L+G FD VPMDSAKSKY+ FLP+ L LL G+++DVPQ G+I
 Sbjct: 121 IQLLEGDAADVLQELKGFPLVPMDSAKSKYVFLPKSELLESENGLLMDVFPQAGEIL 180
 Query: 181 KPIDEVRRGQRTIYRGLQLRLEFSTLQHPDLTATLVPLDGLLMIKKH 227
 PI +++R QR + RSL+ LFD +P TS++PL DGLLMI+K+
 35 Sbjct: 181 LPIMEVRRNRQALRGLRGLRLEFDFVDPNPKYMTSVLPLDGLLMIKKH 227

An alignment of the GAS and GBS proteins is shown below.

Identities = 177/235 (75%), Positives = 199/235 (84%)
 40 Query: 1 MVQSYSTANHNMRPVPVKELVHYMRTQKQKQNGCLAELEAFARQENIPIIHPHETATIF 60
 MV+SYSK ANHNMRPVPVKE+V YMR QKQ G LAELE FA+QENIPII E YF
 Sbjct: 1 MVKSYSTANHNMRPVPVKELVHYMRTQKQITGFLAELEQFARQENIPIIQPEVVAIF 60
 45 Query: 61 RFLMQLSQPKHILEIGTAIGFSALLMAENAPAKITITVTDNRNEMIALAKENFAKYDNNQ 120
 RFL+Q+LQPKHILEIGTAIGFSALLMAENAP+A I TIDRN EMI AK NFAKYD+ Q
 Sbjct: 61 RFLQLSQPKHILEIGTAIGFSALLMAENAPDATTITVTDNRNEMIDPAKANFAKYDSRQ 120
 Query: 121 IRLLEGDAVDVLQTLKSYDFVPMDSAKSKYIVFLPQVLKHLVGGVVLLDIFQGGDIA 180
 I LLEGDA D+L TL+ ++DFVPMDSAKSKYIVFLP++L+ L VGGVV+LDD+FCGGDI
 50 Sbjct: 121 IRLLEGDAADILSTLEGNFDFVPMDSAKSKYIVFLPEILRLLLKVGGVVLLDVPQSGDIT 180
 Query: 181 KPIDEVRRGQRTIYRGLQLRLEFSTLQHPDLTATLVPLDGLLMIKKH 225
 KPI+++RRGQRTIYRGLQ LFD-TL HP+LT +LVPL DGLLMIKKH IVL D
 55 Sbjct: 181 KPIDEVRRGQRTIYRGLQLRLEFSTLQHPDLTATLVPLDGLLMIKKH 225

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2421-

Example 2148

A DNA sequence (GBSx2264) was identified in *S. agalactiae* <SEQ ID 6639> which encodes the amino acid sequence <SEQ ID 6640>. This protein is predicted to be phosphoglycolate phosphatase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 50
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2193 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8985> which encodes amino acid sequence <SEQ ID 8986> was also identified. This protein appears to be a hydrolase *i.e.* an exposed protein.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA91552 GB:267740 unidentified [Streptococcus pneumoniae]
Identities = 39/117 (33%), Positives = 67/117 (56%), Gaps = 9/117 (7%)

Query: 98  KEQESRDSKIHLN-PYAKEILEWTKSQDIPNPMYTHKGASTHSVLETLQISHYFDEILTG 156
20      KE E+R+ + + ++LE Q +P+ +H+ +LE I+ YF E++T
      Sbjct: 25  KENEARLEHSHPLPEGVSDILLELQGRHFLVSHRNDQVLEILEKTSIAAYFTIVVTS 84

Query: 157  VSGFERKHPQGINYLKRRYSLDKSMITYIGDRPLDLEVAQNAGIKS-----INLR 207
      SFG+RCP+P+ + YL ++Y + + IGDRP+D+E Q AG+ + +NLR
25      Sbjct: 85  SSGFKRKPNPESMLYLRKQYQISSGLV--IGDRPIDIEAGQAAGLDPHLFTSIVNLR 139

```

SEQ ID 8986 (GBS240) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 2; MW 26kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 3; MW 51.5kDa).

30 GBS240-GST was purified as shown in Figure 225, lane 12.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2149

35 A DNA sequence (GBSx2265) was identified in *S. agalactiae* <SEQ ID 6641> which encodes the amino acid sequence <SEQ ID 6642>. Analysis of this protein sequence reveals the following:

```

    Possible site: 36
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.2620 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6643> which encodes the amino acid sequence <SEQ ID 6644>. Analysis of this protein sequence reveals the following:

```

    Possible site: 54
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.2967 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-2422-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 463/599 (77%), Positives = 541/599 (90%)

```

5   Query: 1   MSDNRSHIERKYQNDLTTVFATDELWETVEVLTQADNAGFSGHLLDSSQSLLEITEV 60
      M+DNRSRH+ERIKYNDL+T+VATD+ WE EV +L ++ +KGF+GHLLDSS +LL++T+
      Sbjct: 1   MTDNRSHLERKYTNDLSTIFATDKNRAEVDLATEVRAKSGFAGHLLDSSANLLAVTKT 60

10  Query: 61   ELDLSRRLEKVVVYASMKNDQDTTVAKYQEQAKATALYAKPSSETFSFYEPPELLQLSSED 120
      L+L+RR+EKVVVY+ MKNDQDTTVAKYQEQAKA+ LYAKPSE SFY+PE++ L + D
      Sbjct: 61   YLELARRVEKVVVYAHMKNDQDTTVAKYQEQAKASGLYAKPSRVFSFYDEPVMMLHQED 120

15  Query: 121  YQSFLLEMPDLQKYDHFPEKIFANKPRVLVSQNEEELLAGASETPGAAGETFEILDNADMV 180
      YQ+FL E P+L+ Y+HFF+K+F + RVLSQ EEELLAGA EIF A ETF ILDNAD+V
      Sbjct: 121  YQAPLTETPELKVYNHFFDKLPQAREHVLSQAEELLAGAQTIPNGASETFSILDNADIV 180

20  Query: 181  FVVVQARGSEVELTHGNFISLMESSDRTVRKEAYQAMYSTYEQFQHTYAKTLQTNVKSQ 240
      FVVVQEN KGB+VELTHGNFISLMS DR+VR+ AY+AMYSTYEQFQHTYAKTLQTNVQ Q
      Sbjct: 181  FVVVQNDKGEDVELTHGNFISLMSKDRSVRQAAEAMYSTYEQFQHTYAKTLQTNVKSQ 240

25  Query: 241  NFKARVHHYQSARQSALSANFIPERVYETLITVNHHLPLHRYMKLRQKVLGDLDDLY 300
      N+KARVH Y SARQ+A++ANFIE VY+TL++TVN HLPLHRY+KLRQ+VLGDLDDLY
      Sbjct: 241  NYKARVHKYDSARQAAMANFIPERAVYDTLLETVNHHLPLHRYKLKQEVLGDLDDLY 300

30  Query: 301  DVITPLSQMDMSFTYDEALKKSEVLAIIFGRAYSERVRAFTERWIDVHVNKGKRGSGAYS 360
      DVITPLS+ D++ YDEAL+K+E+VLA+FG+ Y++RVHRAFTERWIDVHVNKGKRGSGAYS
      Sbjct: 301  DVITPLSETDLAIGYDEALKAEKVLAVFGKDYADRVHRAFTERWIDVHVNKGKRGSGAYS 360

35  Query: 361  GGSYDTNAPFLNMQDTLDNLVTLVHETGHSLSHTFTRENOQYVYGDYSIFLBIASITN 420
      GGSYDTNAP+LLNMQDTLDNLVTLVHETGHSLSHTFTRE QYVYGDYSIFLBIASITN
      Sbjct: 361  GGSYDTNAPFLNMQDTLDNLVTLVHETGHSLSHTFTRETQYVYGDYSIFLBIASITN 420

40  Query: 421  ENILTEILLKEVKDDKNRFAILNHYLDGFGTIFRQTFAPFSEHAIHVADQSGQVLTSY 480
      ENI+TE LL EV+D+K RFAILNHYLDGF+GT+FRQTFAPFSEHAIH ADQ+G+VLTSEY
      Sbjct: 421  ENIMTEALLINEVQDEKERFAILNHYLDGFRGTVFRQTFAPFSEHAIHQADQKGEVLTSY 480

45  Query: 481  LNNLYAEILNKYYGLTKEDNHFIQENARIPHFYNNYVYQYATGFAAANYLAERIVNGN 540
      LN LYA+LNKYYGL+K+DNHFIQENARIPHFYNNYVYQYATGFAAA+YLA++IV+G
      Sbjct: 481  LNQLYADLNKYYGLSKDNEFIQENARIPHFYNNYVYQYATGFAAASYLADKIVHGT 540

      Query: 541  PEDKAYLYNLYKAGNSDYPLNVIKAGVDMTSADYLDAAFKVFEERLVELEHVARGVH 599
      +D + YL YLK+GNSDYPL VIKAGVDM DYL+AAF+VF+ERL ELE LV+EG+H
      Sbjct: 541  QDDIDHYLAYLKSGNSDYPLNVIKAGVDMKSDYLEAAAFKVFDERLVELEHVSIGIH 599

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2150

A DNA sequence (GBSx2266) was identified in *S. agalactiae* <SEQ ID 6645> which encodes the amino acid sequence <SEQ ID 6646>. This protein is predicted to be competence protein. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.2955 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-2423-

>GP:ABC23745 GB:AP052209 competence protein [Streptococcus pneumoniae]
Identities = 127/269 (47%), Positives = 176/269 (65%), Gaps = 8/269 (2%)

5 Query: 1 MLIADKQGNLINLLESHPGKQYFCPTCCSAVLKAGRIHRHFAHISLANCOFPHENE 60
M +A+D +G L+N+LE K Y CP C + L+ G +R HFAH SLK+C F+ ENH
Sbjct: 1 MFVARDARGELVNVLEDKLEKQAYTICPACGGGLHLKQPSVKTTHFAHSLKDCDFPFENE 60

10 Query: 61 SNEHLQLKAKLYMSLSRENEMLEHHLPEINQIADLVFNETHALE ---VQCRLSEQRL 116
S EHL K LY L +E + LB+ L E QIAD+VFN LALE V C + + L
Sbjct: 61 SPEHLANKPSLYHWLKKCTKVQLLEYPLSELKQIADVFNKGLAESVVVPCLR---KVL 117

15 Query: 117 RERTKAYLQADPQVRWLLGKELMLKHLRLNLHKLQFLQPSQSIGPHIWEIDLRLVELRLKY 176
+ER++ Y +QV WLLG+KLMLK RLT L FL PSQ++GF++WELD +VRLKYI
Sbjct: 118 KERSEGYPSQGYQVWLLGKELMLKHLRLNLHKLQFLQPSQSIGPHIWEIDLRLVELRLKY 177

20 Query: 177 LIYEDLRGHVYTLKTCPL-SGDVLAFLKWPYQSKNMFYKVKQDRNIRDYVROQLRYGN 235
LIY+DLRG ++Y K G +L L+ PY+ + ++ + V +D+I Y+RQQL Y N
Sbjct: 178 LIYQDLRGKLYHYKIFPSYGGSSLELRLPYKQKISHFTVSEDKDICRYTRQQLRYGN 237

25 Query: 236 QFNLRKQKAYISGQNLITQELMFFPQI 264
FW+++Q +AY G+N+L L +PQI
Sbjct: 238 LFWMKQAEAYQKGNLITGLKWEYFQI 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6647> which encodes the amino acid
30 sequence <SEQ ID 6648>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
30 bacterial cytoplasm --- Certainty=0.1034(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 154/312 (49%), Positives = 204/312 (65%), Gaps = 1/312 (0%)

Query: 1 MLIADKQGNLINLLESHPGKQYFCPTCCSAVLKAGRIHRHFAHISLANCOFPHENE 59
+L A D + LI+L+ + K + CP C S VRL+ G I R HFAH+ L +CQF EN
40 Sbjct: 4 ILTALDCKNQLISLVTQPISTKPPFRCPACKSPVRLRQGTIRRHFAHQAHCQFQAN 63

Query: 60 SNEHLQLKAKLYMSLSRENEMLEHHLPEINQIADLVFNETHALEVQCRLSEQRLRER 119
ES EHL LKAKYI SL R +E +LPE+ QIADL+VN+ LALE+QCS L +RL+R
Sbjct: 64 ESPEHLQLKAKLYMSLVRTEAVCTIKYLPQLQIADLVNDELALAIQCSPLPVRSLRKR 123

45 Query: 120 TKAYLQADPQVRWLLGKELMLKHLRLNLHKLQFLQPSQSIGPHIWEIDLRLVELRLKYLI 179
TKAY + + VRMLLG KLML LT L KQFL FS S+GFH+WELD +RLKYLI+
Sbjct: 124 TKAYQKGYVPVRWLLGKELMLKHLRLNLHKLQFLYFSSLGPHIWEIDAAANLRLKYLIH 183

50 Query: 180 EDLRGHVYTLKTCPL-SGDVLAFLKWPYQSKNMFYKVKQDRNIRDYVROQLRYGNQFNL 239
EDL G V YL+KT L +++ + PYQ + L Y+ K N+ +++ L + NL
Sbjct: 184 EDLPKGVSYLTKTISLDHNMIMPELFPYQCSILYSYQKMTVNLKSKRIQALLARHPKWL 243

55 Query: 240 RKQKAYISGQNLITQELMFFPQIOPPRVDTDFQYTNLSLTSFYQNTFYQKQKRNLD 299
R+QKAYISG NLL F+PQ +P + + PQCI +L +Y++F YY+K K+
Sbjct: 244 RKQKAYISGYNLMLTIDAFYQKRPVQSSSGFCQIKGNLRPYRYSFKVYYKKRDKKV 303

Query: 300 QTLYPFVYDKI 311
QTL+ P +Y K+
Sbjct: 304 QTLFSPKYVVM 315

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2151

A DNA sequence (GBSx2267) was identified in *S. agalactiae* <SEQ ID 6649> which encodes the amino acid sequence <SEQ ID 6650>. This protein is predicted to be bicyclomycin resistance protein. Analysis of this protein sequence reveals the following:

```

5      Possible site: 25
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -8.33    Transmembrane    78 - 94 ( 75 - 96)
      INTEGRAL    Likelihood = -8.33    Transmembrane    269 - 285 ( 267 - 287)
      INTEGRAL    Likelihood = -7.38    Transmembrane    290 - 306 ( 287 - 314)
10     INTEGRAL    Likelihood = -7.06    Transmembrane    203 - 219 ( 199 - 225)
      INTEGRAL    Likelihood = -6.69    Transmembrane    157 - 173 ( 143 - 184)
      INTEGRAL    Likelihood = -6.42    Transmembrane    53 - 69 ( 44 - 73)
      INTEGRAL    Likelihood = -6.42    Transmembrane    362 - 378 ( 357 - 381)
      INTEGRAL    Likelihood = -3.72    Transmembrane    242 - 258 ( 240 - 261)
15     INTEGRAL    Likelihood = -3.24    Transmembrane    329 - 345 ( 328 - 346)
      INTEGRAL    Likelihood = -1.28    Transmembrane    107 - 123 ( 106 - 123)

----- Final Results -----
      bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
20     bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

25     >GP:CRA15047.GB:AJ235272 BICYCLOMYCIN RESISTANCE PROTEIN (bcr1)
      [Rickettsia prowasekii]
      Identities = 86/336 (25%), Positives = 159/336 (46%), Gaps = 28/336 (8%)

30     Query: 73 GKKNVLLGLCLLMSGFISFFTSNFSLAMASRLLLGIGIGLYNSLSISIIITDLYEADER 132
      G++ VLIGL ++S IS P+ N + M +R + G+ + + + S+ D Y+ E
      Sbjct: 70 GRRPVLVLLGIYIVSSIIISIFSPNEMLMARFIQAFGVSVGVSGQSMARDYSYGAE 129

35     Query: 133 ASMIGLRTASLNGKALTITFVGLVLA-IGVNYITVLVILVIVPEF-PFKNVPEVENQT 190
      + + + + L AL ++I G + + +Y+++ + L + + + +FE
      Sbjct: 130 SYVYAILSPWLLFIPALGSIYIGYII EYLSMHTVPIFFSLAGITLLALYQILPETNYII 189

40     Query: 191 HTLKASTTFDT-----KAALLMLITFLVGI---AYIGATVKIPTLLVTKHYATS PSSNM 242
      ++S P+ K +L L P++G Y G ++ P +L+ + SF +
      Sbjct: 190 AFSQSSKYFEVFNIIIDKMLWLYAFIIGAFNGIYQFFIAPFLIDQMRVLSPFYGLK 249

45     Query: 243 LTLIAFSGLIIVSVFGKLVK---VFQKLTLLIMLAMGIGNVLPAANNQIIFIVAS--I 297
      LL+P+ I G + G L+K V+ +K + I + G +LFA+ + + Fi+ S
      Sbjct: 250 AFLLSFASIPGGFLGGYLIKRRQYDKKVMISIGFITFSLGCLFAVDSPFLEFIVLNVNF 309

50     Query: 298 LIGASFVGTN-----SSVFYISINAKYKRNNTISLALTAGNI-QVILTFLLI--LTKLP 349
      I F+ M S+ I+ YA E +T TAG+I G I +I +T
      Sbjct: 310 AIAMIFMFMHIMHIGSLIIATILRYALRDYATVVG---TAGSIPGAIYVVIASVYTCV 366

      Query: 350 SOLHLEPFMTPLITISGLMIVNV--FVYLVMSIKK 383
      S+H E L+ L+ +V F Y+ L+ K K
      Sbjct: 367 SKIHGETISNFSLLCLVLSISVSISFYIICLYKK 402

```

A related GBS gene <SEQ ID 8987> and protein <SEQ ID 8988> were also identified. Analysis of this protein sequence reveals the following:

```

55     Lipop: Possible site: -1    Crend: 7
      McG: Discrim Score: 6.28
      QVH: Signal Score (-7.5): -2.45
      Possible site: 25
    >>> Seems to have a cleavable N-term signal seq.
      ALCM program    count: 10 value: -8.33 threshold: 0.0
60     INTEGRAL    Likelihood = -8.33    Transmembrane    78 - 94 ( 75 - 96)
      INTEGRAL    Likelihood = -8.33    Transmembrane    269 - 285 ( 267 - 287)
      INTEGRAL    Likelihood = -7.38    Transmembrane    290 - 306 ( 287 - 314)

```

INTEGRAL	Likelihood = -7.06	Transmembrane	203 - 219 (199 - 225)
INTEGRAL	Likelihood = -6.69	Transmembrane	157 - 173 (143 - 184)
INTEGRAL	Likelihood = -6.42	Transmembrane	53 - 69 (44 - 73)
INTEGRAL	Likelihood = -6.42	Transmembrane	362 - 378 (357 - 381)
INTEGRAL	Likelihood = -3.72	Transmembrane	242 - 258 (240 - 261)
INTEGRAL	Likelihood = -3.24	Transmembrane	329 - 345 (328 - 346)
INTEGRAL	Likelihood = -3.24	Transmembrane	107 - 123 (106 - 123)
PERIPHERAL	Likelihood = 3.71		
modified ALCO score: 2.17			

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4333 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

[illegible]

There is also homology to SEO ID 400

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
60 vaccines or diagnostics.

-2426-

Example 2152

A DNA sequence (GBSx2268) was identified in *S.agalactiae* <SEQ ID 6651> which encodes the amino acid sequence <SEQ ID 6652>. This protein is predicted to be 16S pseudouridylate synthase (rsuA). Analysis of this protein sequence reveals the following:

5 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2645 (Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA306992 GB:AP001518 16S pseudouridylate synthase [Bacillus halodurans]
 Identities = 106/234 (45%), Positives = 141/234 (59%), Gaps = 1/234 (0%)

Query: 1 MRLDKLLGQAGFGSPNQVKKLICSRQSVSDGQIVTKDNVIVDSGLQSI FVGKERVCLKES 60
 MR+DK L GFGSR VKKL+ + V V GQ + + V+ +SI V E V K
 Sbjct: 1 MRLDKFLANNMGFGSRKDVKLLKTKTGA VRVGOPIKDPSTHVEPSESITVYGEERVEYKPY 60

20 Query: 61 SYLLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPYIGRLDRDTTEGLIVTNGPLGY 120
 Y ++ KP GV+ A D EH+TVIDL+ E+++ P+GRLD+DT GLL++TN+G +
 Sbjct: 61 VYIMMNKPKGVICATEDLEHETVIDLLGEERHYEYSPVGRLDKDTVGLGLITNDGKFNH 120

25 Query: 121 RMLHPKHVIAKTYVVEVNGFLERDAITFFEEGVVFDGTECKKPAELTIDTANNKSTARI 180
 ++ PKHEV KTY V G + + + F GVV DGG KPA L I A +S +
 Sbjct: 121 WLMSPKHHVPTKTYRVLVEGHVTEEDVGAFSGHVLDDGGVTKPATLHLEA-GARSHIEL 179

Query: 181 TITBCKFHQVKKMFLAYGVKVIYLRISPGDLRLDNLKPKQYRRLRDSEAIL 234
 +TEGKFHQVK+MF A G +V+ L RI G+L LD L G+YR L E A+L
 30 Sbjct: 180 ILTEGKFHQVKRMFQAVGKRVLELERIKIGLLLDPELARGEYRELTFEETALL 233

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6653> which encodes the amino acid sequence <SEQ ID 6654>. Analysis of this protein sequence reveals the following:

35 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3310 (Affirmative) < succ>
 40 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/194 (57%), Positives = 138/194 (70%)

45 Query: 1 MRLDKLLGQAGFGSPNQVKKLICSRQSVSDGQIVTKDNVIVDSGLQSI FVGKERVCLKES 60
 MRLDKLL GSR+QVKKLI ++ V VD VD GLQ I V +RV +
 Sbjct: 1 MRLDKLGGTKVGRSRQVKKLIAQGVVDHMPARNGRQNVDPGLQLEVTGQEVTHPKH 60

50 Query: 61 SYLLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPYIGRLDRDTTEGLIVTNGPLGY 120
 SY +L KPSGVVSA +D+ + TVID ++E+DK LYP+GRLDRTTEGL++TN+GPLG+
 Sbjct: 61 SYILLNKPSPGVVSAKDDTNYLTVIDLAEEDKSPDLNYPVGRLDRTTEGLVLTNDGPLGF 120

Query: 121 RMLHPKHVIAKTYVVEVNGFLERDAITFFEEGVVFDGTECKKPAELTIDTANNKSTARI 180
 RMLHP HHV+KTY V VNG L DA FF G+ F G +C+PA+LTI A+ D+S A +
 55 Sbjct: 121 RMLHPSHHVSKTYLTVTNGLLAEADSDFFAAGICFPTEGQCPAQLTLKADIDQSQSL 180

Query: 181 TITBCKFHQVKKM 194
 TI+BGKFHQVKK F
 60 Sbjct: 181 TISBGKFHQVKKCF 194

-2427-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2153

A DNA sequence (GBSx2269) was identified in *S. agalactiae* <SEQ ID 6655> which encodes the amino acid sequence <SEQ ID 6656>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9745> which encodes amino acid sequence <SEQ ID 9746> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAAL8872 GB:D90917 hypothetical protein [Synecocystis sp.]
 Identities = 197/318 (61%), Positives = 243/318 (75%)

Query: 22 MGLLVNDGKWDQNYDTASTGGKPFVTVTQFRHWVTGDSAGPSGDAGPKAEGRYHLYVS 81
 MGLLV+G W DQNYDT STGG+PVR +QFRHW+T DGS GP+G GPKAE+GRYHLYVS
 Sbjct: 1 MGLLVNGIWDQNYDTSTGGKPFVTVTQFRHWVTGDSAGPSGDAGPKAEGRYHLYVS 60

Query: 82 LACPWASRVLMIRLKNLSEHISISIVNFMLENGMTPOEYKGVIPDMINQSQQYLYQIYQ 141
 LACPWA R LI RKLK LK I +S+V+ LM ENKMTF GV+PD + +YLYQIY
 Sbjct: 61 LACPWAHRTLIIFRKLKGLKCHIDVSVVHLMRENGMTFAPGPGVMPDPLNARYLYQIYT 120

Query: 142 ASQSDYTGRVTVFVLMKKKPHFTIVNSESSEIMELNATAFNHITGNTDDYYPDSLGQIDE 201
 + + Y+GRVTVF+LNDK+ TIVNSESSEI+R+ N+AF+ + + DYYP +L+ QID
 Sbjct: 121 RADAQYSGRVTVFVLMNDKQKQITIVNSESSEIIRI FNSAFDGLGAKSGDYYPKALRTQIDA 180

Query: 202 MNNFIYPKINNGVYKGFATSONVYQKEVETLFTALDQLEKHLSDNNHVLWGQFTEADIR 261
 +N+ IY INNGVYK GFAT+Q Y++ + LP +LD LE L + YL G++ TEAD R
 Sbjct: 181 LNDRIYHTINNGVYKGFATQQTAYBEAIAPLFESLDNLEGLQGHQILTGEITEADWR 240

Query: 262 LFTTLRFDFVYVGHGFKCNLKLHDIYPHLHRYTKRIYNLPGIASTVNFPHDKKHYYGSHK 321
 LFTTL+RFD VY GHGFKCNL+ + DYP+LW Y + +Y+ PGIASTVNF HIK HYY SH
 Sbjct: 241 LFTTLIRFDVYVGHGFKCNLRRIQDYFNLMRYLRDLVHQPGIASTVNFPHDKKHYYGSHK 300

Query: 322 TINPTGILPAGNLEWTI 339
 INPTGI+P GP LD ++
 Sbjct: 301 NINPTGIVP+GFALELISL 318

No corresponding DNA sequence was identified in *S. pyogenes*.

SEQ ID 6656 (GBS655) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 2-4; MW 27kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2154

A DNA sequence (GBSx2270) was identified in *S. agalactiae* <SEQ ID 6657> which encodes the amino acid sequence <SEQ ID 6658>. Analysis of this protein sequence reveals the following:

Possible site: 43

-2428-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1116(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase
 [Bacillus subtilis]
 Identities = 112/243 (46%), Positives = 163/243 (66%), Gaps = 10/243 (4%)

15 Query: 1 MRVITVKNDIEBGGKIAPFTLLERKMKAGAQ-T-LGLATGSSPITFYEEIVKS----NLDPSN 55
 M++ ++ E K++ +++E++A LGLATGS+P+ Y++++ +DFS
 Sbjct: 1 MKILAEHYEELCKLSAATIKBQIQAKKDVLGLATGSPFVGLYKQLISDYQAGEIDFSK 60

20 Query: 56 MVSINLDEYVGIAASNDQSYSPMHKHLFDKAPFKENNL--PNGLAKDLKESIKRYDAVI 113
 ++ NLDGY G++ S+ QSY++PMH+HLF + ++ P G L+ K Y + I
 Sbjct: 61 VTTNLDGYAGLSPSHFQSYNHFHHEHLFQHINMQPDRHITPQSDNPQLEAACKVYEDLI 120

25 Query: 114 N-ANPDPQLLIGRUGHIGTFNEPQTFDITTHVVLDLPSTIEANSRFFNISID-VFQQA 171
 A ID QILGIG NGHIGFNEPG+ P+ T VV L+ STI+AN+RFF VP+ A
 Sbjct: 121 RQAGGIDVQLIGIGANGHIGTFNEPGSDFEDTRVVVLSESTIQANARFPGDPVLVPEELA 180

30 Query: 172 LSMGIGSIMK-SKTTVLVAVGIEKAKAIAISMIGPITDMFASILQKHDDVVIIVDEAA 230
 +SHGI +IM+ SK IVL+A G EKA+AI M +GP+T D+PASILQKH+ V +I D AA
 Sbjct: 181 ISMGIKTINFEFSKHIVLLASGEKAKDAIQKQABGFVTTDVPASILQKHNVTVIADYKA 240

35 Query: 231 SKL 233
 KL
 Sbjct: 241 QKL 243

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6659> which encodes the amino acid sequence <SEQ ID 6660>. Analysis of this protein sequence reveals the following:

35 Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 174 - 190 (174 - 190)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase
 [Bacillus subtilis]
 Identities = 120/244 (49%), Positives = 162/244 (66%), Gaps = 12/244 (4%)

50 Query: 1 MKIIRVQDQIBGGKIAPFTLLKDSL-AKGAKTLGLATGSSPISFYQEMVKS----PLOSFD 55
 MKI+ + E K++ ++K+ AK LGLATGS+P+ Y++++ +DFS
 Sbjct: 1 MKILAEHYEELCKLSAATIKBQIQAKKDVLGLATGSPFVGLYKQLISDYQAGEIDFSK 60

55 Query: 56 L/SINLDEYVGLSVESDQSYDYFNRQNLF---NAKFPKKNYLPNGLATDVBEAKRYNQI 112
 +T+ NLDEY GLS QSY++PM ++LF N +P ++P G +EA K Y +
 Sbjct: 61 VTTNLDGYAGLSPSHFQSYNHFHHEHLFQHINMQP-DHITPQSDNPQLEAACKVYEDLI 119

60 Query: 113 IAEHP-IDFPVQLGIGRNHIGFNEPGTSPFSETHVVLDQESTIEANSRFFNISID-VFQ 170
 I + ID Q+LGIG NGHIGFNEPG+ FE+ T VV L ESTI+AN+RFF VP+
 Sbjct: 120 IROAGGIDVQLIGIGANGHIGTFNEPGSDFEORTVVVLSESTIQANARFPGDPVLVPEELA 179

Query: 171 AISMGIASIMK-SBNIVLLAPQGEKADAINGVFGPITRHLFASILQKHDIHVIVIDEAA 229
 AISMGI +IM+ S+ IVLLA G+EKADAI+ M GP+T +PASILQKH+HV VI D A
 Sbjct: 180 AISMGIKTINFEFSKHIVLLASGEKAKDAIQKQABGFVTTDVPASILQKHNVTVIADYKA 239

-2429-

Query: 230 ASQ 233

A +L

Sbjct: 240 AQL 243

- 5 An alignment of the GAS and GBS proteins is shown below.

Identities = 163/233 (69%), Positives = 201/233 (85%)

Query: 1 MRVITVNDIEGGKIAFTLLLEEFMKAGAQTLGLATGSSPITPYKEIVKSNLDFSNMVSIN 60
 M++I V++ IEGGKIAFTLL++ + GA+TLGLATGSSPI+PY+E+VKS LDFs++ SIN
 10 Sbjct: 1 MKIIRVQDQIEGGKIAFTLLKDLAKAKTLGLATGSSPISFYQEMKSLDFSLTSIN 60

Query: 61 LDEYVGIAASNDQSYSYFMKHLFDKPKFKNLPLNGLAOLKKEIKRYDAVINANPIDF 120
 LDEYVG++ +DQSY YFM ++LF+AKPKF+N LPMGLA D++ E KRY+ +I +PIDF
 15 Sbjct: 61 LDEYVGLSVESDQSYDYPMRQNLFNKPKFKNYLPLNGLATDVEAEKRYNQIILAEHPIDF 120

Query: 121 QILGIGRNGHIGFNEPGTFPDITTHVVDLAPSTIEANSRFFNSIDVPKQALSMGIGSIM 180
 Q+LGIGRNGHIGFNEPGT F+ THVVDL STIEANSRFF SI+DVPRQA+SMGI SIM
 20 Sbjct: 121 QVLGIGRNGHIGFNEPGTSPFREETHVVDLQESTIEANSRFFTSIDVPKQALSMGIASIM 180

Query: 181 KSKTIVLVAGIEKAEAIASMIKGPITEDMPASILQKHDDVVIIVDEAAASKL 233
 KS+ IVLAA+G EKA+AI M+ GPITE +PASILQKHD V++IVDEAAAS+L
 25 Sbjct: 181 KSRMIVLVAFGQKADAKGMVFGPITEHLASILQKHDRVIVIVDEAAASQL 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 2155

A DNA sequence (GBSx2271) was identified in *S.galactiae* <SEQ ID 6661> which encodes the amino acid sequence <SEQ ID 6662>. Analysis of this protein sequence reveals the following:

Possible site: 61
 30 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.12	Transmembrane	169 - 185 (161 - 194)
INTEGRAL	Likelihood = -6.37	Transmembrane	151 - 167 (145 - 168)
INTEGRAL	Likelihood = -5.15	Transmembrane	42 - 58 (41 - 62)
INTEGRAL	Likelihood = -1.59	Transmembrane	207 - 223 (207 - 224)
35 INTEGRAL	Likelihood = -1.12	Transmembrane	24 - 40 (23 - 40)

----- Final Results -----

bacterial membrane	---	Certainty=0.4248 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAF13747 GB:AF117351 unknown [Zymomonas mobilis]
 Identities = 88/216 (40%), Positives = 123/216 (56%)

45 Query: 9 QQLNLRAGVLGANDGISVAGVVIGVASATHNLMWIFLSAASAILAGAFSMAGGEYVSF 68
 +Q+ LRA VLGNDSI+S ++IGVASA + I L+ S ++AGA SMA GEYVSF
 Sbjct: 17 RQMCNLRASVLGANDGISLSTSSLMIGVASAHSSSGNILLAGMSGLIAGALSMAAGGEYVSF 76

50 Query: 69 STQKITEQAAVAREEKLENNPELAKKSLVDIYIAKESHEHAQWLVDKAFSKNAIRHLV 128
 S+Q D EQA VARE L+ NP K L +IY+ +G EA + ++ + NA+E +
 Sbjct: 77 SSQHMEQADVAREHARKANPAKHEIETAYVERGLRELALQVAVQLMAHNALEAHL 136

Query: 129 EEKYGIEPGEYTSIPWHAIAISSPIAIGSIFPTTITILLPFSVRIVGIVIVISLLSTG 188
 ++ G+ P AA++S I+F+ G+I P +T L P + + +I I+ L G
 55 Sbjct: 137 RDELGLTDSLIRFVQALASATPSGGAIVPPIALFSPPEIINITISLISLCLAVLG 196

Query: 189 YVSAKLQGAFTVPMKRNVMIGLIMLATVYIQQLF 224
 V A LG A A R G L M + T IG F
 60 Sbjct: 197 MVGAHLGGANVPKALRVTFGALAMIGTAAGSF 232

-2430-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2156

- 5 A DNA sequence (GBSx272) was identified in *S.agalactiae* <SEQ ID 6663> which encodes the amino acid sequence <SEQ ID 6664>. This protein is predicted to be S-adenosylmethionine tRNA ribosyltransferase (queA). Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3438 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14732 GB:Z99118 S-adenosylmethionine tRNA ribosyltransferase
(Bacillus subtilis)
Identities = 228/341 (66%), Positives = 279/341 (80%)

20 Query: 1 NNTNDFYFLPEELIAQTPLEKRDASKLLVIDHQNKTMDGSHFDHILDELKPGDALAPNN 60
M + FDF LPS LIAQ PLE+RDAS+L+V+D +TDS F HI+ GD LV+NN
Sbjct: 1 MKVDLDFELPERLIAQVPLEQRDASRLMVLDKHGSELTDSPFKHISFPNKGDCVLVNN 60

25 Query: 61 TRVLPARLYGEXQDTHGRVELLLNNTSGDQNEVLAKPAKRLRVGTQVSPGDRLIATVT 120
TRVLPARL+G K+DT VELLLK GD+WE LAKPAK++ GT V+PDGRL A T
Sbjct: 61 TRVLPARLPGTKEDTGAKVELLLKQBTGDKWETLAKPAKRVKGTVTVTPDGRLAICT 120

30 Query: 121 KELEHGGRIVEFSYDGIPLVLESIGEMPLPPYIHEKLEDRDRTYQTVYKENGSAAPTA 180
+ELSHGR +EP YDGIF EVLESIGEMPLPPYI E+L+D++RYQTVY+KE GSAAAPTA
Sbjct: 121 EELEHGGRKMEFYDGIFFYEVLESIGEMPLPPYIKEQLDDKERYQTVYSKEIGSAAPTA 180

35 Query: 181 GLHFTKELLEKIETKGVKLVYLTLHVGLGTFRPVSVDNLDEHMHSEFYQLSKEAADTLN 240
GLHFT+E+L++++ KGV++ ++TLHVGLGTFRPVS D ++SH MH+EFYQ+S+E A LN
Sbjct: 181 GLHFTKEILLQLKDKGVQIEFTILRVGLGTFRPVSADVEEHNMHAEFYQSESTAAALN 240

40 Query: 241 AVKESGGRIYAVGTTISRTLETIGSKFNGELKADSGWNTIPIKGYQPKVVDVAFSTNHL 300
V+E+GRI++VGTTIS RTLETI + +G+ KA SGWT+IFI PGV+FK +D TNHL
Sbjct: 241 KVRENGGRIISVGTTISRTLETIAGEHDGQFKASGWTISIFTYPGYEFKADGMINFHL 300

Query: 301 PKSTLVMLVSAFAGRDFVLKAYNHAVEERYFFSFGDAMFV 341
PKS+L+MLVSA AGR+ +L AYNHAVEE YRFFSFGDAM +
Sbjct: 301 PKSLIMLVSAAGRENILKAYNHAVEEYRFFSFGDAMLI 341

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6665> which encodes the amino acid sequence <SEQ ID 6666>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3864 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 55 An alignment of the GAS and GBS proteins is shown below.

Identities = 297/341 (87%), Positives = 322/341 (94%)

-2431-

- Query: 1 MNTNDFDFYLPEELIAQTPLEKKRDSKLLVIDHKNKTMDSHFHDHILDELKPGDALVGN 60
 MNTN+PDF LPEELIAQTPLEKKRDSKLL+IDH+ KTM DSHFDHI+D+L PGDALVGN
 Sbjct: 1 MNTNDFDFELPEELIAQTPLEKKRDSKLLITDHRQKTMVDSHPDHIIDQLNPGDALVGN 60
- 5 Query: 61 TRVLEARLYGEEKQDTGHVLELLLNKTEBGDQWEVLAKPAKRLRVGTVKVS PGDGRLIATVT 120
 TRVLEARLYGEEK DTGHVLELLLNKNT+GDQWEVLAKPAKRL+VG++V+PGDGR L AT+
 Sbjct: 61 TRVLEARLYGEEKPDTHGHVLELLLNKNTQGDQWEVLAKPAKRLKVGSGVNPFGDGRLIATII 120
- 10 Query: 121 KELEHGGRIVEFSYDGI FLEVLSELSGSMPLPPYIHEKLEDRDYQTVYAKENGSAAPTA 180
 ELEHGGRIVEFSYDGI FLEVLSELSGSMPLPPYIHEKLESD +RYQTVYAKENGSAAPTA
 Sbjct: 121 DELEHGGRIVEFSYDGI FLEVLSELSGSMPLPPYIHEKLEDAERYQTVYAKENGSAAPTA 180
- Query: 181 GLHPTKELLLEKIETGVKLVYLTLHVGIGTRFVSVVDNLDEHMHSEFYQLSKRAADTN 240
 GLHPT +LL+KIE KGV LVYLTLEHVGIGTRFVSVVDNLDEH+MHSEFY LS+RAA TL
 Sbjct: 181 GLHPTDLLKKIEAGKGVHLYLTLEHVGIGTRFVSVVDNLDEHMHSEFYLSLEBAATLR 240
- 15 Query: 241 AVKSSGGRIVAVGITSIRTLETIGSKFNGELKADSGWTNIPFKPGYQFKVVDAPSTNFHL 300
 VK++GGR+VAVGITSIRTLETIG KF G+++ADSGWTNIPFKPGYQFKVVDAPSTNFHL
 Sbjct: 241 DVKQAGRVAVGTTISIRTLETIGGKFGDQADSGWTNIPFKPGYQFKVVDAPSTNFHL 300
- 20 Query: 301 PKSTLWMLVSAPAGRDVLEAYNHAVEERYFFSFGDAMFY 341
 PKSTLWMLVSAPAGRDVLEAY HAV+3+YRFFSFGDAMFY
 Sbjct: 301 PKSTLWMLVSAPAGRDVLEAYRAH+DEKYRFFSFGDAMFY 341
- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2157

A DNA sequence (GBSx2273) was identified in *S.galactiae* <SEQ ID 6667> which encodes the amino acid sequence <SEQ ID 6668>. Analysis of this protein sequence reveals the following:

- 30 Possible site: 36
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -14.22 Transmembrane 14 - 30 (6 - 34)
- 35 ----- Final Results -----
 bacterial membrane --- Certainty=0.6689 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6669> which encodes the amino acid sequence <SEQ ID 6670>. Analysis of this protein sequence reveals the following:

- Possible site: 51
 >>> Seems to have no N-terminal signal sequence
- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2655 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 50 An alignment of the GAS and GBS proteins is shown below.

Identities = 126/195 (64%), Positives = 155/195 (78%), Gaps = 1/195 (0%)

- Query: 160 MEERFDITETDYKYIGENHNVAAPSGAMSIDMWQKYSLVYSNTPAYALAEIRIGQMDA 219
 M ERFDITETDYEY EH+ YVA P+GAMSI DWQ+YSLVYSNTPAYALAEIR+GGM+ A
 Sbjct: 1 MTERFDITETDYEQSHAYVAQPGAMSIDMWQYSLVYSNTPAYALAEIRIGQMDA 60
- Query: 220 YSKFRYGQSGKSDIKNIQKNGKVTYDIIYQLVDYLWHRKRYDLSLTLYLESAPFTDYR 279
 Y P RYG+ G I I +NGRK+TT YY+QVLDYLW+R KY ++ Y+ E+FP YY+
 Sbjct: 61 YQLFDYRGKYSGLTITDRNGKTIITAYLQVLDYLWHRQKRYDILYYIGESFPDLYK 120

-2432-

- Query: 280 ALIPSDVVVAQKPGYVREALNWGAIVKEVPYVAITAGLGSTQKDSRINGVGLYQLE 339
+P V V QKPGYVREALNWGAIV EE PY++A+Y++GLGG+TQ E+NG+G QL
Sbjct: 121 TYLP-HVKYVQKPGYVREALNWGAIVCEBS FYLIATLYSSGLGGATQASEBVNGLGYQLV 179
- Query: 340 QLCFVINQWHRVNMN 354
QL +VIN+W+R N+N
Sbjct: 180 QLPYVINEWYRGNLN 194
- SEQ ID 6668 (GBS680) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 10-12; MW 64kDa) and in Figure 239 (lane 9; MW 64 kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 15; MW 40kDa) and in Figure 188 (lane 9; MW 40kDa). Purified GBS680-His is shown in Figure 242, lane 8. Purified GBS680-GST is shown in Figure 246, lanes 6 & 7.
- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2158

A DNA sequence (GBSx2274) was identified in *S. agalactiae* <SEQ ID 6671> which encodes the amino acid sequence <SEQ ID 6672>. Analysis of this protein sequence reveals the following:

- Possible site: 17
>>> Seems to have no N-terminal signal sequence
- | | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -4.57 | Transmembrane | 8 - 24 (4 - 25) |
| INTEGRAL | Likelihood = -2.13 | Transmembrane | 66 - 82 (65 - 84) |
| INTEGRAL | Likelihood = -1.65 | Transmembrane | 107 - 123 (107 - 125) |
| INTEGRAL | Likelihood = -0.69 | Transmembrane | 36 - 52 (36 - 52) |
| INTEGRAL | Likelihood = -0.48 | Transmembrane | 89 - 105 (89 - 105) |
- Final Results -----
- >>> bacterial membrane --- Certainty=0.2826 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2159

A DNA sequence (GBSx2275) was identified in *S. agalactiae* <SEQ ID 6673> which encodes the amino acid sequence <SEQ ID 6674>. Analysis of this protein sequence reveals the following:

- Possible site: 59
>>> Seems to have an uncleavable N-term signal seq
- | | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -9.87 | Transmembrane | 108 - 124 (97 - 133) |
| INTEGRAL | Likelihood = -9.08 | Transmembrane | 181 - 197 (173 - 201) |
| INTEGRAL | Likelihood = -7.43 | Transmembrane | 220 - 236 (216 - 248) |
| INTEGRAL | Likelihood = -6.69 | Transmembrane | 6 - 22 (3 - 28) |
| INTEGRAL | Likelihood = -3.72 | Transmembrane | 401 - 417 (400 - 417) |
| INTEGRAL | Likelihood = -3.35 | Transmembrane | 279 - 295 (278 - 295) |
| INTEGRAL | Likelihood = -2.87 | Transmembrane | 31 - 47 (30 - 50) |
| INTEGRAL | Likelihood = -2.87 | Transmembrane | 244 - 260 (242 - 264) |
| INTEGRAL | Likelihood = -0.80 | Transmembrane | 62 - 78 (62 - 78) |
- Final Results -----

-2433-

bacterial membrane --- Certainty=0.4949 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21770 GB:U32694 H. influenzae predicted coding region HI0092

[Haemophilus influenzae Rd]

Identities = 232/416 (55%), Positives = 314/416 (74%), Gaps = 3/416 (0%)

10 Query: 4 TTTTGGALIGLALAILLIKKVHPAYSLILGALVGGIGGDLTVTNMVLGAQGMSS 63
 T + GAL+ L +AI LI+KKV PAY ++GALVGGIGG DL V+ M+ GAQ+ ++
 Sbjct: 3 TVSAIGALVALIVAIFILKKVSPAYGMVLGALVGGIGGADLSQTVSMIGGAQGITTA 62

Query: 64 ILRIILSGILAGALIKTGSARKIASSIIKKLQCCRAITALAIATMIICAVGVFDIAVIT 123
 ++RIL +G+LAG LI++G+A I E+I KIG+ RA+ ALA+ATMI+ AVGVF+D+AVIT
 15 Sbjct: 63 VMRIILAGVLACGLVLSGAANSITETITIKLGSTRALLALALATMLTAVGVFVDVAVIT 122

Query: 124 VAPIALAIGKKANLKSSTLLWMIIGGKAGNIISPNTIAASEAFVLDLSLMQNIIP 183
 V+PIALA+ +++LSK++ILLAMIGGKAGNI+SPNPN IAA++ F + LTS+M+ IIP
 20 Sbjct: 123 VSPIALALSRRLSLKAILLAMIIGGKAGNIMSPNPAIAADTTHPLTSTVMNAGIIP 182

Query: 184 AIALVVTIILAKIVSKONDISYDSEGV--GSDLPAFLPAISGPLVVICLALRPLFG 241
 A+ L+T+ LAK + K + ++ D E V +LP+FL A+ PLV I LALRPLF
 Sbjct: 183 ALFGLILTYFLAKLINGSKVT-DKEVIVLETONLESFLLTALVAPLVAIILLALRPLFD 241

25 Query: 242 ITIDPLIALPGLGISILATGYLKETVPPFVEYGLSKVVGSIILIGTSLSGIKASNLQ 301
 I +DPLIALPGLGI G L+ + GLSK+ V+I+L+GTG L+GII S L+
 Sbjct: 242 IKVDPLIALPGLGIGAPCWGKLRNINSYAINGLSKMFPVAINLLGTG LAGI+IANSGLK 301

30 Query: 302 FDNHLLSFLNMTFIIAPLPSGIFMGAAATSTTSGTTIASQTFAETLIKSGVPAVSGAAM 361
 +I LE +P++IAP+SG+ M ATASTT+GT +AS F+ TL++ GV +++GAM
 Sbjct: 302 EVLIGQLSHSLPSYIIAPISGVLSLATASTAGTAVASNVFSSTLELGVSSLGAAM 361

Query: 362 IHMGATVLDLPHGSPFHATGGAVNMAIKDRMKLISYREALIGTSTIVAVVYCYFF 417
 35 IHAGATV D +PHGSPFHATGG+VNM IK+R+KLI YE+ +GL TIV+ + + F
 Sbjct: 362 IHAGATVDFDMHGPSFPHATGGSVNDIKERKLIPYSAWGLMMTIVTLIPGVF 417

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6675> which encodes the amino acid sequence <SEQ ID 6676>. Analysis of this protein sequence reveals the following:

40 Possible site: 51
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood	~-11.15	Transmembrane	240 - 256 (236 - 265)
INTEGRAL	Likelihood	~-10.88	Transmembrane	3 - 19 (1 - 32)
INTEGRAL	Likelihood	~-10.14	Transmembrane	269 - 285 (263 - 289)
INTEGRAL	Likelihood	~-7.27	Transmembrane	107 - 123 (102 - 141)
INTEGRAL	Likelihood	~-7.17	Transmembrane	307 - 323 (303 - 330)
INTEGRAL	Likelihood	~-6.64	Transmembrane	24 - 40 (23 - 43)
INTEGRAL	Likelihood	~-5.63	Transmembrane	422 - 438 (420 - 442)
INTEGRAL	Likelihood	~-3.77	Transmembrane	124 - 140 (124 - 141)
INTEGRAL	Likelihood	~-3.24	Transmembrane	189 - 205 (184 - 207)
INTEGRAL	Likelihood	~-2.60	Transmembrane	65 - 81 (65 - 82)
INTEGRAL	Likelihood	~-2.34	Transmembrane	393 - 409 (393 - 409)
INTEGRAL	Likelihood	~-0.11	Transmembrane	149 - 165 (149 - 165)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

>GP:BAB07616 GB:AP001520 unknown conserved protein [Bacillus halodurans]

Identities = 155/435 (35%), Positives = 248/435 (56%), Gaps = 21/435 (4%)

Query: 7 LGVLGVGVIVIIIVLVEKNWIIAFLATSLVILFNQMDPTTLLSKRPNFNGALSTYIL 66

-2434-

LG+++G++++ L + +II AP+A +V LF +D LL + +M +
 Sbjct: 2 LGIVLGLVITMVLAYHGSIIWVAPLAAAGVVALPGGID----LLPAYTDTYNEGFVNFPAK 57

5 Query: 67 NYFAIFLLGSSILAKIMETSGAITSADIYLLKKVHDSFYKVLVAIPLISAILTYGGISLF 126
 +P +L+G+I KIME +GA S+A I K +G + + + + L A+LYGGISLF
 Sbjct: 58 QMFVFMGLGAIPLKIMETGAARSVASAITKLIGTK--RAILGVMGLCAVLTYGGISLF 114

10 Query: 127 VVMFAVLPLARSLFKMDLAWNLIQVPLMLGIATPTMILRGTPAIQNVIPQIYDLTSLT 186
 VV+FA+ FLA +LF++ + + LI + LG TETMT +RGT IPQ+IP Y T+
 Sbjct: 115 VVVFAMYLALALFREAMISRRLIGTIALGATPTMTAVNGTPTQIQNLIPSYTYGTAM 174

15 Query: 187 AAIPSIYSGICGVARGLFYMKYCLAKSMARGSETYATYAFDNRIOVKTKNLPHPLASILF 246
 AA + + + + G Y+ + K GE + T + E + + + +P+ S LF
 Sbjct: 175 AAPMNVLAALIMGIGSTYTLVWRKKLKRAG-FFTEPKNGHKEEKEGVFVFNLSFLF 233

20 Query: 247 LLLLIIIALTGSGLPNDFFKNNIIFIALLAIVLTASWLFQPIPNKIAVFNLAGSSSIAP 306
 L+ +I+ T L D I +AL++ I+ L + I N GA S+
 Sbjct: 234 LVSVIV---TLNLQND-----IVLALISGIIVLIMLAVGVKVGFIQSMNQAAGGSVLA 284

25 Query: 307 IPATASAVAGCAVMVIVPQFTFFSDLIILAIQNPILASIVLTSSMSAITGSSSGALIVM 366
 I T++AV FG+VV VFGF ++L+L I G+PLIS AV + + TGS+SG +GI +
 Sbjct: 285 IINTSAAVFGSVVAVVFGFPERLTSLGLGYSPLISCAVINVLGATGSASGAGIAL 444

30 Query: 367 ----PNCAYTLDGCLNPEMIIHRVATIASNIFTIVPQSGVLTPLALGILNKHAFKETE 422
 + Q + + G+FE HRVA+IAS +P +Q LT/LA+XGL+HK ++K+ F
 Sbjct: 345 EALGDRITQLAMETGMSPEAFHRVATIASGGLDTLPHNGAVITLLAITGLSHKESYKDI 404

Query: 423 ITVSVSTFIACVIVI 437
 + V ++ I
 Sbjct: 405 VVGVPIVSVFAFI 419

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/395 (22%), Positives = 167/395 (42%), Gaps = 40/395 (10%)

35 Query: 9 GALIGLALAILLIKKVPAYSILGALVGLIGSGDLVTIV---NTMVLGAQ--MMS 62
 G L+G+ I L +K+V+ + L + L D T + +GA +++
 Sbjct: 8 GVLGVIVIIYLYVKEVNIILAAPLATSLVILFNQMDPTTLTGKSPNGALSTYILN 67

40 Query: 63 SILRIITSGILAGALIKTGSABKIASIIKKLGSQ--RAITALAIATMIICAVGVFIDI 119
 L ILA + +G+ IA+ I+KK+G + + A+ + + I+ G+ + +
 Sbjct: 68 YPAIFLLGSSILAKIMETSGAITSADIYLLKKVHDSFYKVLVAIPLISAILTYGGISLF 127

45 Query: 120 AVITVAPIAIAIGKANLSSKILLAMIGGGKAGNII---SPNPTIAASEAPKVDLTS 175
 + V P+A+ + KK +L+ + I + + G + +P + + LT+
 Sbjct: 128 VVMFAVLPLARSLFKMDLAWNLIQVPLMLGIATPTMILRGTPAIQNVIPQIYDLTSLT 187

50 Query: 176 LMVQNIIPATAALVVITII----LAKIVSKKNDISY--DSEBQVGS-DLPAFLPAISGP 227
 + +I+ +I + + LAK +++ +Y D+E QV + +LP FL +I
 Sbjct: 188 AAIPSIYSGICGVARGLFYMKYCLAKSMARGSETYATYAFDNRIOVKTKNLPHPLASILPL 247

55 Query: 228 LVVICLLALRPLFG-----ITIDPLIALPLGLISILATYKLTGVPPFVYGLSKVVG 280
 L+I+ + LFG I L+A+ L S L ++ + + G S +
 Sbjct: 248 LLLLIIIALTGSGLPNDFFKNNIIFIALLAIVL--TASWLFQPIPNKIAVFNLAGSSSIAP 306

60 Query: 281 ---VSILLIGTGLSIIKASNLQFDMIHLLPLAMPTFILAPLGGIFMGAATSTTSGT 337
 + + G + I+ D+I L P LA L+ M A T S+ +
 Sbjct: 306 IPATASAVAGCAVMVIVPQFTFFSDLI--LNIPLNELISLAVITS-SMSAITGSSSGAL 362

Query: 338 TIASQTFATLLIKSGVPAVSGAMIHAGATVLDL 372
 I FA+ + G+ MIH AT+ ++
 Sbjct: 363 GIVMNFPAQYLYDQSL----NPMIHRVATIASNI 393

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2435-

Example 2160

A DNA sequence (GBS2277) was identified in *S. agalactiae* <SEQ ID 6677> which encodes the amino acid sequence <SEQ ID 6678>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -3.24    Transmembrane    85 - 101 ( 84 - 101)

----- Final Results -----
bacterial membrane --- Certainty=0.2296 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB16041 GB:Z99124 similar to hypothetical proteins [Bacillus subtilis]
Identities = 176/377 (46%), Positives = 234/377 (61%), Gaps = 2/377 (0%)

Query: 1 MKVVAIDSLKGLSSLEAGNAIKESINEVISQADVEVIHPLADGGESTVEALTLGMGGTI 60
MK+AA DS K SLS+LEA AI+ V GAD P+ADGGESTV+LL G I
Sbjct: 1 MKIIAPDSFKESLSALEAAEAIERGFKSVFPGADYRKLPVADGGESTVQSLVDATNGRI 60

Query: 61 ETIPVGLPGEKVHASYGIIIPQRLAIEMAAAGITLATEERNPLHTTYYGVGEMIKD 120
V GPIGE V A +G++ + A+EMAA+G+ L+ ++NPL TTT G GE+I
Sbjct: 61 IEQVVTGIPGEVPVRAFFGMGDGRATAVEMAAASGLHLVVDKRNPLITTRTGELLGA 120

Query: 121 AISKGRHPIITIGGGSATNDGGAGMLQALGYALLDKDNQEISLGAQLADLKSISTDKVI 180
A+ G IIGIGSATNDGGAGM+QALG LLD EI G L+ L SI +
Sbjct: 121 ALDCAERLIIGIGSATNDGGAGMIQALGRLDNGSGEIGPOGALSQLASIDVSGLD 180

Query: 181 EEIKKCDPFKACDVINPLCGAGCCSIFPGQKGADEMITMTWLSNYATLATSVSSEKA 240
L+ ++AC+V NPL G +G +++FGQKGA DM+ +D +S+AA A
Sbjct: 181 SRLRNVLKLEAVCNVDNPLTGPKGATAVFGPQKGAATDMLDVLQNVSHPADMAEKALST 240

Query: 241 DRTIEGTGAAGLGFAFLAFINATLEPGDIIILSEINIEKALSEADLVVTGREGLOGTV 300
EG GAAGGLG+ L + A L+ GIDI-L ++ E + +ADLV+TGGR+D QTV
Sbjct: 241 FRTEGTGAAGGLGWSLLTYLQADLKRGIDIVLEAVDFESIVQDADLVITGGRIDSTV 300

Query: 301 MGKAPIGVAKLAKYGGKVPVAFSGSVTEADILCNQNGIDAFPPVRLISLDEAMSKEVA 360
GK PIGVAK AK Y V+ +GS++ D+ QNGIDA F IV + L++A
Sbjct: 301 HSKTPIGVAKAKSYDYFVIGIAGSISRDNNAVYQNGIDALFSIVPGVPLEDAFEHAAE 360

Query: 361 YQMKETATQVFRLLNL 377
Y M+ TA + I L
Sbjct: 361 Y--MERTARDIAASIKL 375

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6679> which encodes the amino acid sequence <SEQ ID 6680>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -0.27    Transmembrane    360 - 376 ( 360 - 376)

----- Final Results -----
bacterial membrane --- Certainty=0.1107 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AA57927 GB:U18997 ORF_f408 [Escherichia coli]
Identities = 115/345 (33%), Positives = 182/345 (52%), Gaps = 25/345 (7%)

Query: 24 MKILVAIDSKGSGVTSPELNTSVAQALLSVQKLVETRAIDGGESLVALSCTVAGRW 83

```

-2436-

MKI++A DS+K S+++ E+ ++ + + +ADGGBG++ A+ G
 Sbjct: 28 MKITVIAPDSYKESLASEVAQAIEKGFRIFPDQAQYVSVFVADGGSGTVMAMIAATQGA 87
 Query: 84 HQVKTIDLLRRPIKVA--YRHAQAQFIESASIIIGDKITSNSVTAQAISYGLAVKD 141
 L + ++ K AFIS A+ G++ + + TS G G +
 Sbjct: 88 RHAWVTGFLGEKVASMGISGDKTAFTIMAAASGLELVPAKRKDELVTITRSYTGELILQ 147
 Query: 142 AIQKGATQIEIMLGGTGTSDGGKFLSSINDFMT-----GRSYLOTLASPVTLGL 193
 A++ GAT I I +GG+ T+DGG G ++L G L+ZL + + + GL
 Sbjct: 148 ALESQATNIITIGGSATNDGGAGVQALGAALKCDIANGNIGFGGSSNLTL-NDIDISGL 206
 Query: 194 T-----DVTNPFYHGPQGAFAVFGPKGSSLSQIEBTQIASNFAKKVFCQTTI 241
 DVTNP G G + +FGPKG S + I E D S++A+ + + +
 Sbjct: 207 DPLRKDCVIRVACDVINPLVGDNASKIFGPKGASEAMIVELDNLSHYARVIAKKALHV 266
 Query: 242 DLQITPGSGMGGSGAIV-LGGTILTSFGSRIABELNLNLSLQSCDLVTGSCCLDTQS 300
 D++ +PG+GAGG+G A++ LG L SG + LNL+ + C LVITGEG +D+QS
 Sbjct: 267 DVKDVPGAGAGGNGAALMPLAGALKSGTIEVITAIALNEEIHDCITLVITGEGRIDQS 326
 Query: 301 QSGKVPVAIARWAKKYQVPTIALGSSVKIETGAARDFL-AVPSI 344
 GKVP+ +A +AKKY P I + GS+ + G+ + + AVPS+
 Sbjct: 327 IHGKVPVIGVANWAKKYHKPVIGINGSLTDDVWVWHQIGIDAVPSV 371

An alignment of the GAS and GBS proteins is shown below.

Identities = 128/379 (33%), Positives = 194/379 (50%), Gaps = 23/379 (6%)
 Query: 1 MKVVVAIDSLKGLSSLEAGNAIKRSINEVISGADVHVHPLADGSEBTVEALTMGGTI 60
 MK++VAIDS KGS++S E + + + + V +E +ADGGBG++ AL+ + G
 Sbjct: 24 MKILLVAIDSPKGSVTSPLNTSVAQALLSVKQLVETRAIDAGGSSLVALSQTVAGRW 83
 Query: 61 ETIPVKGFLGPKGVHASYGIIPQRQLAIIEMAAAGITLATEERNPLHTTYYGVGEMIKD 120
 + L + +Y + A IE A+ GI I + + T+Y+G +ND
 Sbjct: 84 HQVKTIDLLRRPIKVA--YRHAQAQFIESASIIIGDKITSNSVTAQAISYGLAVKD 141
 Query: 121 AISKGRHFIIIGGSATNDGGAGMLQALGYALLDKINQIEISLGAQLADLKSISTDKVI 180
 AI KG I +GG+ T+DGG G L++L Y + G + L + + + + +
 Sbjct: 142 AIQKGATQIEIMLGGTGTSDGGKFLSSINDFMT-----GRSYLOTLASPVTL 190
 Query: 181 BELKECDPKFIACDVTNPLQAGQCSSIPGPKGADENITKMDTWLSNYATLATSVEKA 240
 L DVTNP G GG ++FGPKG I + D SN+A + +
 Sbjct: 191 LGIT-----DVTNPFHGPQGAFAVFGPKGSSLSQIEBTQIASNFAKKVFCQTTID 242
 Query: 241 DATIBGTGAAGGIGAFAPLAFNTALEPGIDITLSEINTEKAISEADLVTTGEGRLGGTV 300
 TI G+GAGGLG A + TL G I +N+ + + DLV+TGES LD G+
 Sbjct: 243 LQTIPTSGAGAGGLGA-IVLLGGTILTSFGSRIABELNLNLSLQSCDLVTGSCCLDTQS 301
 Query: 301 NGKAPIGVAKLAKKYGKVVAFSGSVTDEALCNQHGIDAFPTVPRLLSLDEAMSKVA 360
 GK P+ +A++AKKY +A GSV + L + + A F I ++ ISL+ A+ K
 Sbjct: 302 SGKVPVAIARWAKKYQVPTIALGSSVKIETGLAARDFL-AVPSIQQPTISLEAIDKITT 360
 Query: 361 YKNKGTATQVFRILINLYN 379
 N+K A + LI +N
 Sbjct: 361 LSNKILAAANMLIAQFN 379

SEQ ID 6678 (GBS409) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 7; MW 45.4kDa).

GBS409-His was purified as shown in Figure 214, lane 6.

GBS409d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 3 & 4; MW 35kDa) and in Figure 188 (lane 12; MW 35kDa). Purified protein is shown in Figure 240, lanes 9-10.

-2437-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2161

A DNA sequence (GBSx2278) was identified in *S.agalactiae* <SEQ ID 6681> which encodes the amino acid sequence <SEQ ID 6682>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1886 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP/AAC21771 GB:U32695 conserved hypothetical protein [Haemophilus influenzae Rd]
Identities = 97/383 (25%), Positives = 175/383 (45%), Gaps = 52/383 (13%)

Query: 1 MGLRQLAQIVTSIKDVCQDINPINFEGIIIPASTNPKRVGEPHEIGLKVAQTCQMIEV 60
M+L K A++IV + +N ++ G-I AS N R + H + + + +E+
Sbjct: 1 MQLDKYTKKIVKRAMKIIHHSVNVMDHGVIIASGNSTRINQRHTGAVLALRNKRVVEI 60

Query: 61 TD---QESYFGTQAGINIPFYNCMLLATIGISGNPNQVQKYALLAQMTRLIKEHE-L 116
Q+ F Q GIN+P +Y + + +GISG P QV +YA L + LI++ L
Sbjct: 61 DQALAQNNPEAQPGINLPIHYLGKNGVGVISGEPTQVQYAEVLVMTAEILVEQQALL 120

Query: 117 DYLDPGRDNEASIVLHHLVSGRELDYIYNQFANQVHLSKTDYRLTFEINSQKQKLL 176
+ + R + + +L L+ LN + + + + +F+H + +L+
Sbjct: 121 EQESWHRRYKEPILQ-----LLHCNLNWKMEQQA--KPPSPDLNKSRRVVLI 167

Query: 177 S-----QSEMSLLNFPDK-----LDTAIYTPNYPNQVWLLSDHMFYYPNI 219
+ +L+N+ ++ LD + + N +LS M
Sbjct: 168 KLLNPAIDNLQNLINYLEQSEPAQDAVILSDQVVLKRWQNS--TVLSAQM-----KT 219

Query: 220 LSKFECEKGLYKVGIGQKSSLSLLK---SYETSILALK-ALWGQKQ--VNIVDDLDLLEL 273
L + K YK+ +G +L L ++ S++++ L LK + + +D+ L+
Sbjct: 220 LLPADYSKQDYKIAGVACLNLPLEBQLPLSPQSAQSTLSYGLKHHPRGIGYVDFEHLRPV 279

Query: 274 LLTSDISNLIKQVNLKALVNL--SENDKIL---LNSVFKHNLSLKECSQELFIHKNTQVYR 329
LL + + + L K L L SE + IL L YF N L ++L+H NT++YR
Sbjct: 280 LLAGLSHSGWQNELIKPLSPFSEKNALYKTLQYVFLSNCIDLYLTAASKLPHVPHNTLRYR 339

Query: 330 LNKIYESTQLNPRNFKDATLLYL 352

LNKI + T L D LYL

Sbjct: 340 LNKIEQITGLFPNKIDKLLYL 362

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2162

A DNA sequence (GBSx2279) was identified in *S.agalactiae* <SEQ ID 6683> which encodes the amino acid sequence <SEQ ID 6684>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0290 (Affirmative) < succ>

-2438-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 5 >GP:AAF89979 GB:AF206272 beta-glucosidase [Streptococcus mutans]
 Identities = 334/475 (70%), Positives = 392/475 (82%), Gaps = 8/475 (1%)
- Query: 4 FPKHFLNGGAVAAQVSGAFRTDGKGLSVQDVLPNGSLG-----FTAKTPTDNLKLE 56
 10 Sbjct: 6 FF++FLWGGG AANQ BGA+ DGKGLSVQDV P G+ K PT DNKLK
 FFEHFLNGGATAAQQFSGAYNQDGKGLSVQDVTPKGQVAQSGSSSLITEKPTEDNLKLV 65
- Query: 57 AIDFYHNYKNDIKLFARMGFKVFRISIAWSRIFPMGDDAPNEAGLQFYDNLDELLKYN 116
 IDFY+ YK DI LFAFMGFKVFR SIAW+RIFPMGDD PNEAGL FVD +FDEL KY+
 15 Sbjct: 66 GIDFYNYRKEDIALFARMGFKVFRISIAWTRIFPMGDDLPNEAGLAFYDKVDFELAKVD 125
- Query: 117 IEPLVTLSHYETPLHLAKTYNGWADRLLIAFFKFAQTMERYKDKVKYMLTFNEVNSIL 176
 IEPLVTLSHYETPLHLA+ YNGWA+R LIAP+E++A+TV RYKDKVKYMLTFNEVNS+L
 Sbjct: 126 IEPLVTLSHYETPLHLAKTYNGWANRLLIAFYERYAKTVFTRYKDKVKYMLTFNEVNSVL 185
- Query: 177 HMPFTSGALMTDKSQLSPQELYQAIHHELVASARVTKLGRSINPNFKIGOMILAMPAYPM 236
 H PF SG I+TD QLS Q+LYQA+HHELV SA TK+G INP+FKIGOM+LAMPAYPM
 20 Sbjct: 186 HAPFMGGGIIIDPQLSKQDLYQAVHHELVVSALATKVGHEINPDPFKIGOMVLAAMPAYPM 245
- Query: 237 TSDPRDVLAAQFQBQNLFLSDIHVRGKYPTTYIQSYFRNNGIKIKPEBGDEEVLQAQNTVD 296
 T+DP D LA R+FE N LFSD+H RGKPY YI+ YFK+N I K BGD+E++ +NTVD
 25 Sbjct: 246 TADFLQDLAVREFENQYLFSDLAHARGKYPYIKYFKDNNIDIKWGBGDEELMLTPTVD 305
- Query: 297 FLSFSYMSVTQAYDFENYQSGQGNILGGLTNPLHTISEMOWQIDPIGLRLVINYQYERY 356
 F+SFSYMSV A++ E+Y SG+GN+LGLL+NF+L SEMOWQIDP+GLRLVIM Y+RY
 30 Sbjct: 306 FISFSYMSVAARHPEDYNSGRGNVYGLLSNPLYQASEMOWQIDPVGLRLVINDYERY 365
- Query: 357 QIPLFIVENGLGAKDQLIETLDGDTYVEDDYRIDYMMQHLVQAKAIEDGVIMGYTSMG 416
 Q+PLFIVENGLGARD L++ DG TVEDDYRIDY+ +HL+QV +A++DGV++GTY+MG
 Sbjct: 366 QIPLFIVENGLGAKDVLVQDGP- TVEDDYRIDYLQGLHMLVQVGAQLQDGLDLYTVMG 424
- Query: 417 CIDCVSMSTAQLSKRYGLIYVDRNDDGTSLGKYKKKSPGWYQVKIKNGSLPE 471
 ID VS ST +LSKRYG IYV NDDG+GSL RYKKKSF WY+KVI+TNG SL+E
 35 Sbjct: 425 PIDLVSESTVLSKRYGYFYVACNDDGSGSLARYKKKSPFWYKVIETNGSLYE 479

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5287> which encodes the amino acid sequence <SEQ ID 5288>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0763 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 50 An alignment of the GAS and GBS proteins is shown below.

Identities = 390/469 (83%), Positives = 423/469 (90%)

- Query: 1 MTFPPKHFLNGGAVAAQVSGAFRTDGKGLSVQDVLPNGSLGDTAKTPTDNLKLEAIDF 60
 55 Sbjct: 1 M+FPK FLWGGGAVAAQVSGAFD KGLSVQDVLPNGSLG+T PT DNL LEAIDF
 NG:FPKFLWGGGAVAAQVSGAFDAGKGLSVQDVLPNGSLGEMTDSPTSDNLLEAIDF 60
- Query: 61 YINYKNDIKLFARMGFKVFRISIAWSRIFPMGDDAPNEAGLQFYDNLDELLKYNIEPL 120
 YH YK DI LFAFMGFKVFRISIAWSRIFPMGDD PNEAGLQFYD+LDELL Y IEPL
 Sbjct: 61 YHRYKEDIALFARMGFKVFRISIAWSRIFPMGDDQDNEAGLQFYDNLDELLKYNIEPL 120
- Query: 121 VTLSHYETPLHLAKTYNGWADRLLIAFFKFAQTMERYKDKVKYMLTFNEVNSILHMPF 180
 VTLSHYETPLHLAK TYNG DRLLI FFE+FAQTMERYKDKVKYMLTFNEVNSILHMPF
 60 Sbjct: 121 VTLSHYETPLHLAKTYNGWADRLLIGFFERFAQTMERYKDKVKYMLTFNEVNSILHMPF 180

-2439-

Query: 181 TSGAINTDKSQLSQELVQAIIHRLVASARVTKLGRSINPNFKIGCMILAMPAYMTSDP 240
 TSG IMT+K +LS Q+LQAIIHRLVASA VTKL IMP+ K+SCMILAMPAYMTSDP
 Sbjct: 181 TSGGIMTEKGLSLQDLFQAIIHRLVASASVTKLAHEINPDVKGCMILAMPAYMTSDP 240

5 Query: 241 RDVLAARQFEQGNILLFSDIHVRGKYPTTYIQSYFQNGIKI KFESEGEDESLAQITVDFLSP 300
 RD+LAA FE NLLFSDIHVRGKYP+YI+SYFK NQI+I FE+GD+E+LA++VDFLSP
 Sbjct: 241 RDIILAAHAFENMLLFSDIHVRGKYPYIKGKYKNGIIEIVFEDGCKELLAHEVDFLSP 300

10 Query: 301 SYTMSVTQAYDFENYQSQGNILGSLNPHLITSHQWQIDPIGLRLVLAQYRYVQIPL 360
 SYTMSVTQA++ E Y SQGNILQGL+NP+L +SEWQWQIDPIGLRLVLAQYRYVQIPL
 Sbjct: 301 SYTMSVTQAHNFEATYTSQGNILGSLNPHYLESSEWQWQIDPIGLRLVLAQYRYVQIPL 360

Query: 361 FIVENGLGAKDQLIEITLDDYTVEDDYRIDYIMQHLVQAKAIEDGVIRIMSYTSMGCIDC 420
 FIVENGLGAKDQL++T DG TV DDYRIDYH+QHLVQAKAIEDGVE+MSTYSWGCIDC
 15 Sbjct: 361 FIVENGLGAKDQLVQTDGSMIVHDDYRIDYMSQHLVQAKAIEDGVEVMSTYSWGCIDC 420

Query: 421 VSMSTAQLSKRYGLITVDRNDGTSGLQRYKKKSGFWYQKVITNQSSL 469
 VSMSTAQLSKRYG IIVDRNDGSG L RYKKKSF WY++YI+TNG+ L
 20 Sbjct: 421 VSMSTAQLSKRYGFIIVDRNDGSGTQUTRYKKKSFWDYRQVITNQSSL 469

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2163

A DNA sequence (GBSx2280) was identified in *S. agalactiae* <SEQ ID 6685> which encodes the amino acid sequence <SEQ ID 6686>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood	-10.40	Transmembrane	247 - 263	(241 - 273)
INTEGRAL	Likelihood	- 9.55	Transmembrane	429 - 445	(424 - 450)
INTEGRAL	Likelihood	- 4.88	Transmembrane	285 - 301	(280 - 303)
INTEGRAL	Likelihood	- 3.82	Transmembrane	207 - 223	(205 - 225)
INTEGRAL	Likelihood	- 3.40	Transmembrane	113 - 129	(112 - 139)
INTEGRAL	Likelihood	- 1.97	Transmembrane	309 - 325	(305 - 328)
INTEGRAL	Likelihood	- 1.59	Transmembrane	395 - 411	(395 - 411)
INTEGRAL	Likelihood	- 1.49	Transmembrane	174 - 190	(173 - 193)

35

----- Final Results -----

bacterial membrane	---	Certainty=0.5161(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAAB4286 GB:Z34526 beta-glucoside permease [Bacillus subtilis]
 Identities = 225/594 (37%), Positives = 351/594 (58%), Gaps = 11/594 (1%)

45 Query: 4 YQRTAKAILAAYGGKRNQIHVTHCVIRLRLLDNDEIVNDQVIKTIKPNVGVWRRNDYQ 63
 Y + +K IL VGGH+N+Q V HC+TREL L ++ + ++ +P V+G + +Q+Q
 Sbjct: 3 YDKLSKDIILQLVGGHNVQVQIHVCHMTRELFLNHLHRAKADRSQLEQLGVNMTNISGQFQ 62

50 Query: 64 IILGNVDVNNYNAFLALGHFENTREPSQKKSSILKLIETIAGVITPLIPALLGGGM 123
 II+GNDV Y A + + + GS +K +L + + I+GV TP++EA+ G GH+
 Sbjct: 63 IILGNVDVFKYQAVIRHSLNDEKSGAGSSSQKWL SAVFVIGSVFTPLIPALDAGGM 122

55 Query: 124 KVIIGLLPMLGIASSSSQVAFINFGDAAYFPMIDMAYSAARFKVTPVIAATVGILL 183
 K + L G + SQ + GD A+YE+P++A SAA +F P +AA + +
 Sbjct: 123 KSLVALAVTFQWAEKSGVHVILTAVGDGAPYFLPLLLMSSARKKSGSNPYVAALANAL 182

Query: 184 LHPAFVTVABKGLPLAGCAPVTLASVSSVIPILIMVFMQYIKRWINKIVPSVUNKSL 243
 LHP ++ GKP+S G PVT A+Y S+VILP+ ++ Y+R+WI+ + + +
 60 Sbjct: 183 LHPDLTALLGAGKPISIGLFPVTAATYSSTVPIILSIWIASYVERNDRPHASLKIV 242

Query: 244 QPTLIIISGLFALVVVGVGLGVIIKGLSSAMLSIYHVAWPLALSILGAMPLVVMTH 303

-2440-

PT +LI L L+ VGPLG I+G+ LSS + ++ A +A+ +L L++MGMH
 Sbjct: 243 VPTFTLLIVVPLTLLIVGPIGALLGSLVSSGVNVLFDHAGLVAMILLAGTFSLLIMTMH 302

5 Query: 304 WAFAPFLAASVATPVLLPAMLASNLQAQASLAVAVAKQKQKTRQVAFAGLSALLA 363
 +AF PI + +LPM +N+ Q AS AV +++ K+ + A+ +A+L+
 Sbjct: 303 YAFVFMININIAQNGHDYLLPAMFLANMQGASGVFVLRSSNKKFKSLALTTSTALM- 361

10 Query: 364 GITEPALYGVTLKFKKPLYAAMLSGLLVGAYIGLVNIASTYTFVPSYIIGLPQYINPQGN 423
 GITEPA+YGV ++ KKP AA+I G GA+ G+ +ASV +V GLP I G
 Sbjct: 362 GITEPAMYGVMRLKPKFAAALLGGAAGGAFYGMGVASY- -IVGNGAGLPS-IPVFIGP 418

15 Query: 424 NFSNAVIAAIAITILLFTIITWFLGIDGSRNEKSSINAGETHIRSGLSKKEKTYLSPMVG 483
 F A+I + + LG ++ ++ S Q H S +E ++SE+ G
 Sbjct: 419 TFIYAMIGLVIAFAANSTAAAYLLGFDVPSDGSQ---QAVHEGS---REIHSPIKGE 471

20 Query: 484 VLPLSKVFDETFSSKLLGGLALITPSVGEVYAFDGEIISLFFTKHALAKODKGVEVLI 543
 V LS+V D FS+ ++G+G AI P GEV +P G + ++P TKHAI + D+G E+LI
 Sbjct: 472 VKALSEVKGDFVSAGVMGKGLFLEPSGEVVSFVRGSVTITFKTKHAIIGTISDQGAELI 531

Query: 544 HIGIDTVELNGEGFEQLVKVGDVFKKQQLLRMDIDPISNGYSLISPVVVTNS 597
 HIG+DTV+L G F +K GD V G L+ D++ I + GY +I+PV+VTN+
 Sbjct: 532 HIGIDTVKLGQWPTAHKEGCKVAPGDFLVSFDLEQIKAGYDVITPVIVTNT 585

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2883> which encodes the amino acid sequence <SEQ ID 2884>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -10.40 Transmembrane 246 - 262 (240 - 271)
 INTEGRAL Likelihood = -6.26 Transmembrane 284 - 300 (279 - 304)
 INTEGRAL Likelihood = -4.14 Transmembrane 173 - 189 (172 - 194)
 INTEGRAL Likelihood = -3.24 Transmembrane 112 - 128 (111 - 137)
 INTEGRAL Likelihood = -2.39 Transmembrane 428 - 444 (425 - 445)
 INTEGRAL Likelihood = -2.13 Transmembrane 383 - 399 (380 - 401)
 INTEGRAL Likelihood = -1.97 Transmembrane 308 - 324 (304 - 327)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40

An alignment of the GAS and GBS proteins is shown below.

Identities = 508/619 (82%), Positives = 561/619 (90%), Gaps = 1/619 (0%)

45 Query: 4 YQETAKAILAAVGGKNIQGVTHCVTRLRVLNDRIVNDQVIKTIPIVIGVMRKNQYQ 63
 YQETAKAILAAVGG+ NIQ VTHCVTRLRVL NDE V DQ +K I NVIGVMRKN QYQ
 Sbjct: 3 YQETAKAILAAVGGKTNIQGVTHCVTRLRVLNDEKVDQVKAINSVIGVMRKNQYQ 62

50 Query: 64 IILNDVNNYNAFLALGHFENTREFPSQKSSILKLIETTAGVITPLIPALLOGGML 123
 IILNDVNNY AFL+LGHF+N + SS+ K SILE+LITTAGVITPLIPALLOGGML
 Sbjct: 63 IILNDVNNYQAFLSLGHFNDQEDHSKANGSILRLIETTAGVITPLIPALLOGGML 122

55 Query: 124 KVIGILLPMLGIASSSSQTVAPINFPDAAIYFPMIAYSAARFKVTVPLAATVGILL 183
 KV+GILLPMLG+AS+ SQTVAIPNFPDAAIYFMP+MIAYSA+RKVTVPLAAT+ GILL
 Sbjct: 123 KVVGILLPMLGLASDSQTVAPINFPDAAIYFPMIAYSAARFKVTVPLAATAGILL 182

60 Query: 184 LHPAFVFMVABGKPLSLGAPVTLASYGSSVIPILMVPLMQYIERWINKIVPSVMSFL 243
 LHPAFV MVAEGKPL+LGPAPVT ASYGSSVITL+MV+LMQYIE+N+N++VPSVMSFL
 Sbjct: 183 LHPAFVFMVABGKPLTLGAPVTVASYGSSVIPILMVPLMQYIERWINKIVPSVMSFL 242

65 Query: 244 QPTLLILISGFALVVVGPLGVITIGKGLSAMLGIYHVAFWLALILGAIMPLVVTGWH 303
 QPTLLILISGFALVVVGPLGVITIG+GLS+ ML+IYHVAFWLAL+ILGAIMPLVVTGWH
 Sbjct: 243 QPTLLILISGFALVVVGPLGVITIGGLSNTMLAIYHVAFWLALILGAIMPLVVTGWH 302

Query: 304 WAFAPFLAASVATPVLLPAMLASNLQAQASLAVAVAKQKQKTRQVAFAGLSALLA 363
 WAFAPFLAASVATPVLLPAMLASNLQAQASLAVA K KQKQKTRQVA AAG+SALLA

-2441-

Sbjct: 303 WAFAPIFLAASVATPDVILILPAMLASNLQAASLAVAFKTKQCTROVALAAGISALLA 362

Query: 364 GITEPALYGVTLKPKKPLVYAMISGGLVGAIGLVNIASTYFVPSIIGLPQYNQCGN 423
GITEPALYGVTLKPKKPLVYAMISGGLVGA+IG VNIASITYFVPSIIGLPQYNP G3

5 Sbjct: 363 GITEPALYGVTLKPKKPLVYAMISGGLVGAFIGVNIASITYFVPSIIGLPQYNPSS3A 422

Query: 424 NFSNAVIAAATITILITFIITWFLGIDGENEKSSINAGBHTHIRSGLSKKEKTLTSPMVG 483
NF+NA+IA ATIL L F +TWPF+GIDE E+ K A + + ++SGLS K+TLY+PM G

10 Sbjct: 423 NPTNALLAGTATITVLAFSLTWPMFGIDE -SPKQVSVAAAMSCVKSGLSTKQTLVAFMTGR 481

Query: 484 VLPLSKVPDETFSKLLGSGLAITPSVGEKVYAPFDGRIISLFFPKHAIKAKDDGVEVLI 543
+L Ls+VPDETFSKLLGSG AI PS GEVYAPFDGE+I+ FFKHA+ALK+ +GVEVLI

Sbjct: 482 MLPLSKVPDETFSKLLGSGFAILLPSGEKVYAPFDGEVITFFPKHAVALKNTGVEVLI 541

15 Query: 544 HIGIDTVELNGSGFELQVKGDFVKRGQILLRMDIDFISRGYSLISPVVVTNSIDQLI 603
H+GIDTVEL G+GFELV VGD VKRGQ LL+MDIDFI+SRGYSLISPVVVTNS +QLI

Sbjct: 542 HVGIDTVELRGQGFELVSGDVVKRGQALLRMDIDFISRGYSLISPVVVTNSAQLI 601

Query: 604 IVKDAETMVNEDOLLVIL 622
I++D + MYT ED LLVIL

20 Sbjct: 602 IIQDKKMWIKEDALLVIL 620

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2164

A DNA sequence (GBSx2281) was identified in *S. agalactiae* <SEQ ID 6687> which encodes the amino acid sequence <SEQ ID 6688>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1148 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15944 GB:Z99124 transcriptional antiterminator (BglG family)
[Bacillus subtilis]

40 Identities = 118/275 (42%), Positives = 183/275 (65%)

Query: 1 MIKKRVININAVISVTHQGLDVLIMGKGIAPKKRIGRIDNSDAIEKSFVLKNSDNNEFT 60
M I +V+N+N + V Q3 +++++MG+G+AP+K+ GD ++ IER F L N D +F

Sbjct: 1 MKIAKVININAVISVTHQGLDVLIMGKGIAPKKRIGRIDNSDAIEKSFVLKNSDNNEFT 60

45 Query: 61 ELFTITVPEVVACSERIINLGKIKLGNLDEILYINLTDHHSATERHEBQGVQINLRL 120
L +P E + SE II+ K++LGR L++ +Y++LTDHI+ AI+R+++G+ I+N L

Sbjct: 61 TLTYDITPEVNESEEIHYAKLQKGLKNDISYVSLTDHINPAIKRQKGLDKNALM 120

50 Query: 121 EIQRYYPDEYSIGMKALELIDELGICLTIDBSAPIAMIEFVWNLGNPNPEAHKITEIVS 180
E +R Y DE++IG +AL ++K++ G+ L DE+ FIA+H VNA L+ IT+++

Sbjct: 121 ETKRLKYDEFAIGKEALVWKNITGVSLPDEAGFALHIVNRELNERPMNI INITKVMQ 180

Query: 181 YIEQKVKIDPRTLEDESSIDYFRMTHIKLFAQRVLGSMKYVEDDDALLLVKKKPYREY 240
I VK F+ E +E S+ YRFP+TH K FAQR+ +G E D LL VK+KY R Y

55 Sbjct: 181 EILSTVKYHFKIEPNESLHAYRFTHLKFFAQLRFLNGTHMESQDFLLDTVKCKHPRAY 240

Query: 241 KCVKRIKGNMAIQYQYQANSESLLYLTIVHVKIAVK 275
+C K+I + +Y+++L S ELLYLT+H++R+VK

60 Sbjct: 241 ECTKRIQTYIRREYEHKLTDSLELYLTITIERVVK 275

-2442-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6689> which encodes the amino acid sequence <SEQ ID 6690>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0680 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 220/279 (78%), Positives = 246/279 (87%)

Query: 1  MLIKRVLNHNNAVIVTHQSLDVLIMGKGIAPKRIQDRINSDAIEKSFVLKNSDNWNRPT 60
N+IKRVLNHNA IS HQGLD+LLMGEGI P K+GD I +AIE SFVLKNSDNWNRPT
Sbjct: 1  MLIKRVLNHNAAISTHQSGLDILLMGKGITPGKVKGDSIELNAIETSFVLKNSDNWNRPT 60

Query: 61  ELFITVPEEVVACSERIINLKGKIKLGNLDEILYINLTDHIHSAIERHQQMVQINPLR 120
ELFITVP+EVVACSERIINLKGKIKLQ LDEILYINLTDHIHSAIERHQQM+I NPLR
Sbjct: 61  ELFITVPEEVVACSERIINLKGKIKLGTLDLDEILYINLTDHIHSAIERHQQMLHNPLRW 120

Query: 121  EIQRYYFDEYSIGMKALELIKDELGICLTIDESAFIAMHFVNAGLNPFNEAHKITEVS 180
EIQRYYFDEYS+G+KALELI+ LG+ L IDE+AFIAMHFVNA LD PF E H+TEIVS
Sbjct: 121  EIQRYYFDEYSIGVKALELIERNLGVTLAIDEAAFIAMHFVNASLTPPFKEPHRLTEIVS 180

Query: 181  YIEQKVKIDFTELDRSSIDYYRPMTHIKLFAQRVLSGMKGYRDDADLLLVVKIKPKREY 240
YIEQK+K DP+TELD+SIDYYRPMTH KLFACQVLS M Y+DDDA+LLLVVK KPY+EY
Sbjct: 181  YIEQKIKTDFKELDQTSIDYYRPMTHIKLFAQRVLSQMSYRDDADLLLVVKIKPKREY 240

Query: 241  KCVLQKIGNMAIQOVQINSSSELLVLTVHVKRLVNLKE 279
+CV +I + +Y Y LNSSELLVLTVHVKRLVNLKE
Sbjct: 241  RCVLDISEEIKKRYNHLNSSLSELLVLTVHVKRLVNLKE 279

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2165

A DNA sequence (GBSx2282) was identified in *S.agalactiae* <SEQ ID 6691> which encodes the amino acid sequence <SEQ ID 6692>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1104 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9335> which encodes amino acid sequence <SEQ ID 9336> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6693> which encodes the amino acid sequence <SEQ ID 6694>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3314 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-2443-

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/178 (80%), Positives = 161/178 (90%)

```

5   Query: 1   MTLRHDKHHTTYVANANAALKEHPEIGEDLEALLADVSIQIPEDIRQAVINNGGHLNHAL 60
      MTLRHDKHHTTYVAN NAALKEHPEIGR+LE LLADV++IPEDIRQ +INNGGHLNHAL
      Sbjct: 24 MTLRHDKHHTTYVANTNAALKEHPEIGENLEELLADVTKIPEDIRQTLINNGGHLNHAL 83

10  Query: 61  FWEELMSPERTQISQRLSEDINATPGSPGDFKAAFPAAATGRPGSGNAWLVVNARGKLEVL 120
      FWEEL+SFE+ ++ ++++ I+ PGSP+ FK PTAATGRPGSGNAWLVVN EG+LE+
      Sbjct: 84 FWEELSEPKQDVTPDVAQIDDAFGSPDAFKBOFTAATGRPGSGNAWLVVNKEGQLEIT 143

15  Query: 121 STANQOTPMIBGKKPILGLDVWEHAYILYRNVRPNYIKAFPEIINMNKVNELYQAQK 178
      STANQOTPI EGKKPIL LDVWEHAYILYRNVRPNYIKAFPEI+NW KV+ELYQAQK
      Sbjct: 144 STANQOTPISEGKKPILALDVWEHAYILYRNVRPNYIKAFPEIVNWKVSELYQAQK 201

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2166

- 20 A DNA sequence (GBSx2283) was identified in *S.agalactiae* <SEQ ID 6695> which encodes the amino acid sequence <SEQ ID 6696>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

```

- 25 ----- Final Results -----
- | | |
|-------------------------|--|
| bacterial cytoplasm --- | Certainty=0.3331 (Affirmative) < succ> |
| bacterial membrane --- | Certainty=0.0000 (Not Clear) < succ> |
| bacterial outside --- | Certainty=0.0000 (Not Clear) < succ> |

- 30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2167

- 35 A DNA sequence (GBSx2284) was identified in *S.agalactiae* <SEQ ID 6697> which encodes the amino acid sequence <SEQ ID 6698>. This protein is predicted to be DNA polymerase III delta subunit. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

```

- 40 ----- Final Results -----
- | | |
|-------------------------|--|
| bacterial cytoplasm --- | Certainty=0.0511 (Affirmative) < succ> |
| bacterial membrane --- | Certainty=0.0000 (Not Clear) < succ> |
| bacterial outside --- | Certainty=0.0000 (Not Clear) < succ> |

- 45 A related GBS nucleic acid sequence <SEQ ID 9743> which encodes amino acid sequence <SEQ ID 9744> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6699> which encodes the amino acid sequence <SEQ ID 6700>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 43

-2444-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.22 Transmembrane 250 - 266 (249 - 266)

----- Final Results -----

bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 222/340 (65%), Positives = 262/340 (82%)

Query: 1 NIAIKRIGRITPDNLGLVTVLAGEDLGQYAKMKRKLQVIGPNKDDLAYSYFDLSEEDYQ 60
 MIAIE+I +++ +MLGL+T++ G+D+GQY+Q+K +L + I F+KDDLAYSYFD+SE YQ
 Sbjct: 1 MIAIEKIRKLSKENLGLTLVTGDDIGQYSQLKSRIMBQIAFDKDDLAYSYFDMSNAAYQ 60

Query: 61 NAEZDLSESLPFLSDYKVVIFDQFCDITTDKXTYLDEQAMKRFAYLQNFVTTTLVICA 120
 +AE+DL SLFF ++ KVVIFD DITT+KK++L E+ +K FEAYL+NF++TTTL+I AP
 Sbjct: 61 DAEMDLVSLFFPABQKVVITPHLLDITTNKSLFKERDLKAFAYLENPLSTTLIIIFAP 120

Query: 121 GKLDGKRRLVKKLRDARVLEANTLKESDLKTYFQKYAQHGBLVFEAGVDELLIKSNYD 180
 GKLD KRRVLKLLKRD VLEAN LKE++L+TYFQKY+HQ GL FE+G FD+LL+KSN D
 Sbjct: 121 GKLDGKRRLVKKLRDLVLEANPLKFAELRTYFQYSHQGLGPFESGAFDQLLLKSNDD 180

Query: 181 PSDTLNIAFLSKYKTGDHISSNDVREAIKSLQDNIPDLTQDVLGRIDLARDLVRDLR 240
 FS +N+AFLK+YK G+IS D+ +AIPKSLQDNIPDLT+ VL G+ID ARDL+ DLR
 Sbjct: 181 FSGIMQMAFLKAYKKTGNISLTDIEQAIPKSLQDNIPDLTRLVLGGKDAARDLHDLR 240

Query: 241 LQGEDIKLIAIMLGQFRMLQVKILASKGKSESQIVSELSHYIGRKINPYQVKFAVRDS 300
 L GED+IKLIAIMLGQFR+FLQ+ ILA K+E Q+V IS +GR++NFPYQVK+A++DS
 Sbjct: 241 LSGEDIKLIAIMLGQFRFLQLTTLARDVINEQQVLISLSDILGRRVNPYQVKYALKDS 300

Query: 301 RNLPIAFLKEAIRILTIETDYAIKRGTYDKOYLFDLALLKI 340
 R L LAFL A++ LIETDY IK G Y K L D+ALLKI
 Sbjct: 301 RTLSLAFLGAVKTLIETDYQIKTGLYEKSYLVLDIALLKI 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2168

A DNA sequence (GBSx2285) was identified in *S.agalactiae* <SEQ ID 6701> which encodes the amino acid sequence <SEQ ID 6702>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3071(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2169

A DNA sequence (GBSx2286) was identified in *S.agalactiae* <SEQ ID 6703> which encodes the amino acid sequence <SEQ ID 6704>. This protein is predicted to be esterase. Analysis of this protein sequence reveals the following:

-2445-

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -0.32 Transmembrane 175 - 191 (175 - 191)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1128(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB17013 GB:L38252 esterase [Acinetobacter lwoffii]

Identities = 63/218 (28%), Positives = 107/218 (48%), Gaps = 3/218 (1%)

15 Query: 105 KVIFYVHGGSYIHQASELYIFVNKLAKKLDAKVFFPIYKAPTNYSDAIPKIKLYQN 164
 ++IF++HGG++ + + LA + +V+ YP AP + Y +AI I +YQ
 Sbjct: 73 QLIFPHIHGGAFPLGSINTHRALMTDLAARTQMCHVHVVDYPLAPEHPYPEAIDAFDVYQA 132

20 Query: 165 TLASVTSRKQIILVGSAGGGLAIGLADNLVTHIKQPKIILISPLWDIATNNPKIEKV 224
 L PK II+ G+S G LAL L L + P +IL+SP+LD+ + +
 Sbjct: 133 LILVQGIKPKDIIISGDSGANLALALCLRLKQPSLMPSGLILMSPYLDLTLTSGSLRPN 192

25 Query: 225 QKQDPLLKAMQLQQVAFYHANGKINFRNPQVSPLYSSQFNKMAPIFFGITHDIFYPDNQ 284
 QK D LL LQ ++ +P+VSL+ + + P +G+ +I D++
 Sbjct: 193 QKHDAIISIALQAGIKHYLTDDICQDFRVSPLF-DLDGLPPTLVQVGSKEILLDDSK 251

Query: 285 LLHQKLAKENIKHHYIVGQRNMHVYPLP--IPEAETA 320
 +K + ++K H+ + M H + + PEA+ A
 Sbjct: 252 RFRKAEQADVKVHFLYTGMMNINFMFNAMFFPAKQA 289

30 There is also homology to SEQ ID 3498.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2170

A DNA sequence (GBSx2287) was identified in *S.agalactiae* <SEQ ID 6705> which encodes the amino acid sequence <SEQ ID 6706>. This protein is predicted to be purine nucleotide synthesis repressor. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2970(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16124 GB:Z99124 similar to transcriptional regulator [LacI family] [Bacillus subtilis]

Identities = 111/300 (37%), Positives = 175/300 (58%), Gaps = 4/300 (1%)

50 Query: 1 MTSISDIACKAGVAKSTVSRVINHHHPVSDTRQKVMALITELDYIPNQALDLSRGKTQ 60
 M+ I ++LA+ A V+ STVSRV+NHHP+VS+E R+ V ++ ELDY PN+ A DL RGT
 Sbjct: 1 MANIKETARLANVSVSTVSRVLNHHHPVSEKKRLVHCVMKELDYTPMRTAIDLINGKTH 60

55 Query: 61 KIGVPIPTHTPPYTQLINGLLDAKTTIYQLVMMPSDYNQELSLSYLQKLMBAIDALI 120
 +GV++P++ HP F ++NG+ AA +Y ++P++YN ++E+ YL+ L+ + ID LI
 Sbjct: 61 TVGVILPYSNHCPEKIVGVITGAAPGHEYATLLPTNYNPDIEIKYELLEKTKIDGLI 120

Query: 121 FTSRAISLDIIRTYAKYGRIVVCEKLYNHLSSAYLDROYSSFLEAFSDMKLGLGHEHLVL 180
 TSRA D I Y +YQ ++ CE + + + A+ DR +++ E+F +K RG E++

-2446-

Sbjct: 121 ITSRAHWDSSILAYQRYGVPIACEZTGDDID-VPCAPNRDKTAYAESFRYLKSRGHEINIAF 179

Query: 181 LFSRNNESSATYQSALLAYQEVYQOLSSPYMVGVNVDNFNDG-LNLSYQLVKEVSIDGIL 239
 R + S + AY+ V G+L +M+ G +D NDG L + + I

5 Sbjct: 180 TCVREADRSPSTADKAAAYKAVCGRLDRHMLSG-CNDMNDKELAAEHFYMSPGVTPTAY 238

Query: 240 ATSDEVAAGLIKGYRESRKKCPYIIQGECLVQQLKLEPTIDHKSYYLGLKLPKQALARK 299
 A SDEVAAG I + + IIG+ + ++L P++D LG AF L ++

10 Sbjct: 239 ANSDEVAAG-IHLFAKKNRWDEIIGBNTSISRVLGFPFSLDLNLBQLGLIAFSLFLQDE 297

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2171

A DNA sequence (GBSx2288) was identified in *S.agalactiae* <SEQ ID 6707> which encodes the amino acid sequence <SEQ ID 6708>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3451 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AA221682 GB:U32686 conserved hypothetical protein [Haemophilus influenzae Rd]
 Identities = 79/264 (29%), Positives = 134/264 (49%), Gaps = 16/264 (6%)

Query: 1 MTKRIFCDMDGTLINSEGGVSKSNATLIREAA---IPVTLVSARAFEMKKAVDALQLG 57
 M K F D + GTLL S + +S +I+ IF +SAR+P+ + L+

30 Sbjct: 1 MYKAVFSDFNGTLLTSQHTISPRTVVVKRLTANGIFVFPISARSPLGLLFWKQLETN 60

Query: 58 GQVAFNGGLIYRIGDNNQVLPPIHTQLIKKSTVKQLLRGIRHFHPFQVSLSYDLANWYCD 117
 V VAF+G LI N + PI++ I+ + ++ + H P + ++YY N+ +

35 Sbjct: 61 NVLVAFSGALIL----NQNLFEIYSVQIEPKDILEINTVLAEH-FLLGVMYYTNDCHRR 115

Query: 118 KID-EGIRYEHSLTQQCPTFIHNEDQFLBHGINTFKIMMITFDKANMLEKYLQSLLEP 176
 ++ + + YE S+T+ IH D+ T + + I + ++E L+ + P

Sbjct: 116 DVENIGVYRSVTK---IEIHPFDEVA---TRSHKIQIGEAEBIIEIVLLEK-KPF 168

40 Query: 177 EYTIQRSGKAYLEITHLLAKKSGIAYILQKEQLAREETAAPGDGHNDLWMLNMGVPIV 236
 ++I RS +LE+ H A K + ++ + E AFGD NDL MLE VG +

Sbjct: 169 HLSICRSHANFLFVHKSNATKGSVAVRFLDYFGVGTNEVIAFGDNFNDLMLRHVGLGVA 228

45 Query: 237 MDNAFDKIAIKAYQLTKSNDEGDV 260
 M NA ++IK A +T +N+EDG+

Sbjct: 229 MGNAPHEIKQAANVTATNNDGL 252

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2172

A DNA sequence (GBSx2289) was identified in *S.agalactiae* <SEQ ID 6709> which encodes the amino acid sequence <SEQ ID 6710>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.2854 (Affirmative) < succ>

-2447-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2173

- 10 A DNA sequence (GBSx2290) was identified in *S.agalactiae* <SEQ ID 6711> which encodes the amino acid sequence <SEQ ID 6712>. Analysis of this protein sequence reveals the following:

Possible site: 40

- >>> Seems to have an uncleavable N-term signal seq
- | | | | | |
|----|----------|---------------------|---------------|------------------------|
| 15 | INTEGRAL | Likelihood = -10.51 | Transmembrane | 392 - 408 (376 - 417) |
| | INTEGRAL | Likelihood = -9.92 | Transmembrane | 440 - 456 (433 - 461) |
| | INTEGRAL | Likelihood = -6.42 | Transmembrane | 52 - 68 (51 - 70) |
| | INTEGRAL | Likelihood = -6.32 | Transmembrane | 29 - 45 (9 - 48) |
| | INTEGRAL | Likelihood = -6.32 | Transmembrane | 309 - 325 (308 - 328) |
| | INTEGRAL | Likelihood = -4.46 | Transmembrane | 12 - 28 (9 - 29) |
| 20 | INTEGRAL | Likelihood = -3.29 | Transmembrane | 463 - 479 (462 - 479) |
| | INTEGRAL | Likelihood = -2.07 | Transmembrane | 353 - 369 (352 - 369) |
| | INTEGRAL | Likelihood = -1.17 | Transmembrane | 374 - 390 (374 - 390) |
| | INTEGRAL | Likelihood = -0.85 | Transmembrane | 247 - 263 (247 - 263) |
| | INTEGRAL | Likelihood = -0.06 | Transmembrane | 278 - 294 (278 - 294) |
- 25 ----- Final Results -----
- bacterial membrane --- Certainty=0.5203 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23742 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 325/705 (46%), Positives = 478/705 (67%), Gaps = 3/705 (0%)

- 35 Query: 1 MLQLTKYFFPKPIYLLAVFQIYLLVFSWMLGCAFLFSFIPLIYQYDRETIKIAIV 60
 MLQ K F + IYL+ L+ +Y +FS + L +F + L Q+ ++ K + I
 Sbjct: 1 MLQWIKNSFIPLIYLSFILLNLYIAFISASYLALLGVFLLVCLFQFPWKSAGKVLIC 60
- Query: 61 IFFLFYFLWQNNHNNVQYQVFNHISQIKVRIDITISINGDLVSFOADASGNTQYAFYTLK 120
 F F P+Q+N + Q + + + + + DT+ +NGD LSP+ A G +Q +Y L+
 40 Sbjct: 61 GIFGFVWFQWQSQSQNLADSVVERILPVTVKNGDSLSPFGKADRIQVYVYKQ 120
- Query: 121 NKSEKDYFQNLNNIMILADIKLEAEERRRHFGPDFYQVYLRKHGIYRIAKVTKIKQIRL 180
 ++ EK+ FQ L + I + KL R E +R+ F G+Y+ YLK GIY+ + K+ ++
 45 Sbjct: 121 SEEEKFAFQALTDLHEIGLEGLSEFEGQRNFQGNQYQVYLTQGIYQITNKKIQSLQK 180
- Query: 181 FQHRFFALMSKWRSAIVISQT-FENPMRHYMSGLLPGVLDKTFODMSDLYSSIGI IHL 239
 +S RR A+V +T FP+PMR+YH+GLL G+LD F+M+L+LYSSIGI IHL
 Sbjct: 181 IGSWDIGENLSSLRIGVWVWIKTHFPDMRMYMGLLGLDGLDTFEMNELYSSIGI IHL 240
- 50 Query: 240 FALSGMQVGFFIGIPYICIRIGRLRDLHVMQLQIPPSLIYAGLTGPSISVVRALQSLLS 299
 FALSGMQVGFF+ F+ + LR+GL + + L PFSLIYAGLTGPS SV+R+L+Q LL+
 Sbjct: 241 FALSGMQVGFFMNGFKLLRIGLTQEKLLWLYTFFSLIYAGLTGPSASVIRSLQKLLA 300
- 55 Query: 300 HSGVKKIDENFALCLLICISLPSHLLTGGVLSFAYAFILMTWSFOHFSIKKVAIESLT 359
 GVK +N AL +L+ I +P+ T GGVL S AYAFILM S + +K VA ESL
 Sbjct: 301 QHGVKGLDNCALTIVLPLTVMNPFPTAGVLSFAYAFILTMPSEKG-EGLKVAESLIV 359
- Query: 360 VSVGILPILTYFSGFQPSIILTALLSPAFDII FLELLTIVFVLSPIVKLSCLNSLFSI 419
 +S+GILPIL++YF+ FQ P SI+LT + SP PD+ FLELL++FVLS + + +N +FE

-2448-

Sbjct: 360 ISLGILFSLSPFYFAEPQFWSILLTFVFSFLFDITFLPLLSILFVLSFLYPVQIQLNFIWE 419

Query: 420 LEVLLKWTGQLFPPPLIFGKPSLELLIIMLILGLLYDYYHSKCFRYCSLLIIFLTPFIT 479
LE +++ Q+ RPL+FG+P+ +LLI++I L L+YD + L+I LF +T

5 Sbjct: 420 LBGIIRLVSQVTSRPLVFGQPTWLLILLISLALVYDLRKNKIKLIVLCILLTGFLFIT 479

Query: 480 KNPITNEVAILDVGGQDSILVRDMLGKTLITDTGRVR -FPQPEWIKQKVNQSNKAKRTLI 538
K+P+ NE+ +LDVGGQ+SI +RD GKILID GG+ +++ ++W++K+ SNA+R+LI

10 Sbjct: 480 KHPLENEITMLDVGGQESIFLADVTKITLIDVGGKAEYSKKIKKWKQKMTTSNARQSLI 539

Query: 539 FYLKSREGISKIDDLVITHTIDHMKDMEVISKHFKVARLITSKGLINSQVNVHLSKIGV 596
FYLKSREG++KID L++T+TD +H+GD+ ++K F V ++ S SL ++V L

Sbjct: 540 FYLKSREGVAKIDQLILITNTDKHVGDLSPMTKAFHVGEILVSKDLSLQKEFVDELQATQT 599

15 Query: 599 AVKSIKAGDKLAVMSYQLVLYPMHKKDKKNDSDIVLYGHLLKGFLETKDEEKGKQL 656
V+S+ G+ L + GS L+VL P GDG ++D++VLYG L K FLFTG+LEE+GK L

Sbjct: 600 KVRSMIVGENLPIFGQLEVLSPFKNGDGHDDTLVLYGKFLDKQFLFTGNLEEKGEKL 659

20 Query: 659 LEAYPNLSVDILKAGHSGKSSGLSFLKXKLSPSVVLVSAGKNR 703
L+ YP+L V++LGA HG+K SSS +FL+KL P + L+S GK+NR

Sbjct: 660 LKRYPDLKYNVLKAGHSGKSSGFALEKLPKELITLISVSKNR 704

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6713> which encodes the amino acid sequence <SEQ ID 6714>. Analysis of this protein sequence reveals the following:

25 Possible site: 29
>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.19	Transmembrane	394 - 410 (380 - 422)
INTEGRAL	Likelihood = -8.28	Transmembrane	54 - 70 (52 - 72)
INTEGRAL	Likelihood = -6.32	Transmembrane	356 - 372 (355 - 377)
INTEGRAL	Likelihood = -4.73	Transmembrane	8 - 24 (7 - 25)
INTEGRAL	Likelihood = -4.30	Transmembrane	30 - 46 (29 - 50)
INTEGRAL	Likelihood = -3.88	Transmembrane	249 - 265 (249 - 267)
INTEGRAL	Likelihood = -3.40	Transmembrane	467 - 483 (465 - 484)
INTEGRAL	Likelihood = -2.39	Transmembrane	325 - 341 (325 - 347)
INTEGRAL	Likelihood = -0.43	Transmembrane	441 - 457 (441 - 458)

35

----- Final Results -----

bacterial membrane	---	Certainty=0.5076 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

40

The protein has homology with the following sequences in the databases:

>GP:AC23742 GB:AP052208 competence protein [Streptococcus pneumoniae]
Identities = 311/706 (44%), Positives = 456/706 (64%), Gaps = 10/706 (1%)

45 Query: 5 WTKLVPLSKIQFAPLILVFPYQIHSPLWFL-LSLSLICILVKRLSKK--EPLGVFAL 61
W K + I +FL+L +Y I S S+L L L+CL ++ K + L+ I

Sbjct: 4 WIKNFSIPLIYLSFILLMLYIAFSASYIALIGFVFLVLCFLIQFPWKSAGKVLIOGIF 63

50 Query: 62 SPFALFLYKQKQVLQKLEIQVQITSVALVPEIRINDQVLAVRHGKHSYOLFYRLK 121
F +F +Q+ Q Q L + V +PD+++NGD L+ G+ +Q+Y+L+

Sbjct: 64 GPWFVFNWQSQSQASQNLADS---VERVRLPDVTKVNGDSLSFGKAGXGRIFQVYKQLQ 120

55 Query: 122 SQAEALQPKKHKHRLVMHAKVLEKAEVRNPEGNQYFLTYQGIYRIGKVBQIEQLEV 181
S+ E+ F+ + + L+ E+ RNP GNWY +L QGIV+ ++I+ I+

Sbjct: 121 SEEEKAPQALTDLHEIGLEBGLSEPEQGRNFGGPNYQATLKQGIYQTLMIKKIQSLQK 180

Query: 182 ISPESICDYLSSLRRAIVHCQGHFPMWSHYLTGLLWYLDKSGEMIDTYYSOLGIHL 241
I I + LSLRR+R+V + HFF PM +Y+TGLL G+LD F IM + YS LGIHL

60 Sbjct: 181 IGSMDIGENLSSLRRAVWIKTHFPDMNMYTGLLGHLLTDPEMMNELYSLSGIHL 240

Query: 242 FALSGMQGVFPFLTCFRVRLVLLVLEWIKWIEFPACFYAALTNGYSIVIRSLVQSLR 301
FALSGMQGVFP+ P+++LL L + E+K+ PF+ YA LTG+S SVIRSL+Q L

Sbjct: 241 FALSGMQGVFPFMRGPKKLLRLGLTQKRLMLTYPPSLIYAGLTGFSASVIRSLQKLLA 300

65

-2449-

Query: 302 HLGKGLNLA CTFLVFLMDAHLMTVGGVLTFSYAFILTVVTVZELSGAKRQLQVLVT 361
G+KGLN A T L+P+ +E T GGVLA +YAF+LT+ +E G K + L
Sbjct: 301 QHGKGLNLA CTFLVFLMTVMENFFTAGGVLSYAFILTMP6-KHGKGLKAVASESLV 359

5 Query: 362 ISLGILFLFLFYSSNNHSMVLTGLSLFLDLFLFLCLVFLCLSPVTVSICNHLFLIL 421
ISLGIL L FYF+ P P S+LT + S-LDFL LPL+ +F IS L V N +P
Sbjct: 360 ISLGILFLFLFYFARFQNSILLTFVFSFLDLFTPLLSILFLVSLFLYFVTVQNFPIEW 419

10 Query: 422 LKVIQFLGNTFNSSILVPSPTSMHLLILVISFAIFDYRQ-VRQVITCGLVIALTALS 480
LE +L+ + + LVPG P +M L+L+IS A YD R+ ++ + C L+ L L+
Sbjct: 420 LEGIILVSGVTSRFLVFPQNTWLLILLISLAVYDLAKNKIKLTVLCLLITGLFLIT 479

15 Query: 481 VKYPLTNVETPIDIQQDSILVRWETGKNLLIDVQGR-PFSSKEHWRRGHVANAQKTL 539
K+FL NE+T +D+QGS-SI +R+ TKG +LIDVGA+ + +W+ +NA+L
Sbjct: 480 -KHEPLNSTITMLDVQGESIFLRDVTGKTLIDVQGRKASTKKIKWQKMTTNAQRSL 538

20 Query: 540 IPYLSKRGINTIDQLLVITADTHMGDIEVAKAIRIKEILTSQGSLSHPVRLRLK 599
IPYLSRG+ IDQL+T+ D +H+CD+ +KA + EIL S+ SL FV L+ +
Sbjct: 539 IPYLSKRGAKIDQLLITNTDKEHVGDLSEMTKAFHVGEILVSKDSIAQKRFVRLQATQ 598

Query: 600 CHVRVLAAGDQLPMSVQLVLYPQGLDGGIINSLVYKRLNRTFLPTGDLKEKGENE 659
VR + G+ LPI GS L+VL P +GDG +D+LVYNG L+ FLFTG+LE+G+R +
Sbjct: 599 TKVRSMIVGENLPIFGSQLVLSPRKMGDGGHDDTLVLYGKFLDKQFLPQNLLEEGEKD 658

25 Query: 660 IIKRYPOLRVVDYLKAGHNGSTSSAAPLNDTOPKVAPTSAGKNNR 705
+K YD L+V+ LKA HG+ SSS AFL+ +P+ IS GK+NR
Sbjct: 659 LKHYPLKAVVLKASQKNGKSSAPLEKLKPELTLISVGKSNR 704

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 346/743 (46%), Positives = 491/743 (65%), Gaps = 3/743 (0%)

Query: 5 TKYFLKPIYLALLVFIYQLVFSWIMLGCALLESIFLIYQYDRETIKTAIVIFFL 64
TK PL I A L+ + + S+L L L+ + + AI+P
Sbjct: 6 TKLVLFSKIQAFILILVFFYQIHSPSWLTFILSLSLCLNVLKSLKGLVGFALFSCA 65

35 Query: 65 FYPLNQNNHNVQYQYRVNHHISCIKVRIDTISINGVLSQADASGNTYQAFITLANKSE 124
+L+Q + + + F I+ + + D+I INGD L+ ++YQ FY LK+++E
Sbjct: 66 LFLLYKQQLVQKLEIQVQITSVALVPDSIRINGDQLVGRGHSHYQLFLYLSQMS 125

40 Query: 125 KDYPQLNINIMITADIKLSAEKRRHFNQFDYRKYKLEHGIYRIAKVTIKQIRLFQHR 184
F+ +++ A + LS+AE R+F GF+Y+ +L GIYRI KV +I+Q+ +
Sbjct: 126 AQLFKKRRHVLVMAKVTLSKAEVRNFGFNQYQTFILTYQIYRIKGVQIQLEVISPE 185

45 Query: 185 SFFALMSKWRSAIV-ISQTFPNMRHMYSGLLFGYLDKTFDMSDLYSLGIIHLFALS 243
S +S RR AIV Q FP FM HY+GLLFGYLDK+F +M+D YS LGIIHLFALS
Sbjct: 186 SICDYLSSLRRAIVHCOGRFFRPMHSYLTLLFGYLDKSGFMDTYTSQSIIHLFALS 245

Query: 244 GMQVGFPLSIFRYICLRIQRLDRHVVLLQIPFSLIYAGLTQFSSVVRALITQSLSHSQ 303
GMQVGFPL FR + L+ + +L+ + +++FP+ YA LIG+SISV+R+L+QS L H G+
Sbjct: 246 GMQVGFPLTCFRKRVLLLAFLVLEWIKWLELFPACFYAALTQYSIVRSIVGSLRLHGI 305

50 Query: 304 KNDENFALCLILCLISLEHSLITQGVLSYAFYAFILMTWTSFDHPSSIKKVAIBELVSVG 363
K +N A L+ + H L+T GGVLA+YAF+LT+ + + S K+ ++ LT+S+G
Sbjct: 306 KGLNLA CTFLVFLMDAHLMTVGGVLTFSYAFILTVVTVZELSGAKRQLQVLITISLG 365

55 Query: 364 ILPLITYFSGFQPIBIIITALLSFAFDIIFLITLTVIFVLSPIVLSLSCINSLELLEVL 423
ILE L +YFS F P+S+LT LLS+ FD+ LPL+ +F LSP+V +S N LP +E+
Sbjct: 366 ILPLITYFSGSNPMGMVLTLLSYLFDLFLFLCLVFLCLSPVTVSICNHLFLILEK 425

60 Query: 424 LKWTQGLFPRPLIKGKSLGLILVMIITLGLLNDVYHSC-PRYCSILLIPLTFPIFTSN 482
+++ G F L+PG P+ +L++I+ +YDY + C L+I TL + K P
Sbjct: 426 IQFLGNTFNSSILVPSPTSMHLLILVISFAIFDYRQVRQVITCGLVIALTILSV-KYP 484

65 Query: 483 ITNEVAILDVQGDLSILVRWLGKTTLIDTQGRVFPQZHEWKKQKNGSNAKRTILPYLK 542
+TNEV +D+QGSILVR+W GK +LID QER F E+W+ + +NA++TLPYLK
Sbjct: 485 ITNEVETIDIQQDSILVRWETGKNLLIDVQGRPFSSKEHWRRGHVANAQKTLIPYLK 544

-2450-

Query: 543 SRGISKIDDLVITHTDTHMGDMVEISKHPKVARLITSGSLTNSQVGHLSKIGVAVKS 602
 SRGI ID L+TH DTDHMGD+EV+K ++ ++TS GSL++ +V+ L++ V+
 Sbjct: 545 SRGIHTIDQLLVTHADTDHMGDIRVVAKAIIRIKKILTSQSSLSHPSPVRLRLRLKCHVRV 604

5 Query: 603 IRAGDKLAWGSSYLQVLYFWHKGDKGKNDISVLYGHLLGKGFPLPTGDLSEEGEKQLLEAY 662
 + AGD+L +MS LQVLYFW GDGKNDIS+VLYG LL + FLPTGDL+HGE ++++ Y
 Sbjct: 605 LAAGDQLPIMSGVLYQVLYFWQLGDKGKNDISVLYGRLLNRTPLPTGDLSEEGENEITIKRY 664

10 Query: 663 PNI.SVDILKAGHHGSKSSSLSPFKILSPSVLVSAQNNRYQHPHOETLQRPQIKSKI 722
 P L VD LKAGHHGS 888 +PL + P V +SAGNNRYQHPH+KIT. R. + +
 Sbjct: 665 PQLRVLYLKAGHHGSSVSSAFLDHQPKVAFISAGNNRYQHPHIRETLARLEDROITY 724

Query: 723 PRTDQSGTIRLIGWGHIIQTVR 745
 +RTD G IRLIG WH++TVR
 15 Sbjct: 725 YRTDTQGAIRLTGRTSWHLETVR 747

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2174

- 20 A DNA sequence (GBSx2291) was identified in *S.agalactiae* <SEQ ID 6715> which encodes the amino acid sequence <SEQ ID 6716>. This protein is predicted to be competence protein (comEA). Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq
 25 INTEGRAL Likelihood = -3.77 Transmembrane 18 - 34 (14 - 36)

----- Final Results -----
 bacterial membrane --- Certainty=0.2508 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]
 35 Identities = 96/217 (44%), Positives = 138/217 (63%), Gaps = 4/217 (1%)

Query: 3 EIVLEKIKSHKMETTGIIVGILLFGILGLNHFQ-THKEDNLNINLEK-KVSTITEKKVP 60
 E ++EKIK +K +GLL+ G L T KR NL + ++EK+V
 Sbjct: 2 EAITEKIKKEYKLIIVICTGLLGVGFFLLKPAQCFVKEITLQAEVAVSKDLVSEKVN 61

40 Query: 61 NISHVKKVSNQVITVDVKGAVNHFQVYSLPSQSRVTDALKRAGGLSNLADSKSVNLQKL 120
 + ++TVDVKGAV PG+Y LP SR+ DA+++AGGL+ ADSKS-NLAK+
 Sbjct: 62 KKEKEEPLEQDLITVDVKGAVKSPGIYDLFVGSRLNDVQKAGGLTQADSKSVNLQKQV 121

45 Query: 121 QDETVIYVQKGEKITVVERKANNIATQNSKSGKINLMKADLSSLTQISGVGAKRAQDI 180
 DE ++YV KGE+ V ++ A+ + + K+NIMKA L L+ + G+G KRAQDI
 Sbjct: 122 SDERALVYVFTKGE--AVSQQTGLGTASSISKEKKVNHINIKASLEELQVKGGLGKRAQDI 179

Query: 181 LDYRDSQGGPKTIDDLKNVSGIGKTELEKLRQDVITD 217
 +D+R++ G FK++D+LK VSGIG KT+EKL+ VT+D
 50 Sbjct: 180 IDHRSANGKFKSVDRILKVSIGIGKTELEKLDYVITVD 216

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6717> which encodes the amino acid sequence <SEQ ID 6718>. Analysis of this protein sequence reveals the following:

Possible site: 36
 55 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.61 Transmembrane 22 - 38 (16 - 42)

----- Final Results -----

-2451-

bacterial membrane --- Certainty=0.4843 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]
 Identities = 82/179 (45%), Positives = 124/179 (68%), Gaps = 4/179 (2%)

10 Query: 42 NRQSKAAMPALREISPVQKQVSEKKIQLQEDSSILVDLKGAVQKGVYKLTASSRVRDVI 101
 N Q+ A + + + + K+ ESK+E E I VD+KGAV+ G+Y L SR+ D +
 Sbjct: 42 NMQASVAAVS-KDLVSEKRVNKEKESPLEQDLITVDVKGAVKSPGIYDLPVGSRLNDIV 100

Query: 102 ELAAGLTSEADKHAINFAKLTDEQVYVVPKQGEISVLPRSLVSGKKETASKDQSKVHI 161
 + AGGLT +AD ++N A+K++DE +VYVP +GEE + + G + SK++ KV++
 15 Sbjct: 101 QKAGGLTEQADSKSLNLAQKVSEALVYVPTGEE--AVSQQTGLTASSIKKEK-KVNL 157

Query: 162 NKASLEELQHIFPGIGAKRAQDIIDMRDKLGPFKALEDLRQVSGIGKTEKLKDDIFLD 220
 NKASLEEL+ + G+G KRAQDIID R+ G FK+++L++VSGIG KT+EKLKD + +D
 20 Sbjct: 158 NKASLEELKQVKGKGRKRAQDIIDHREANGKFKSEVDELKVKVSGIGKTEKLDYVTV 216

An alignment of the GAS and GBS proteins is shown below.

Identities = 81/166 (48%), Positives = 111/166 (66%), Gaps = 10/166 (6%)

25 Query: 62 ISHVKDKVSNQ-----VTVDVKGAVNHPGVSYLSPQSRVTDIAIKRAGGLSNLADSK 112
 IS VK +VS + + VD+KGAV GYV L + SRV D I+ AGGL++ AD
 Sbjct: 55 ISPVKQVSEKKEIQLQEDSSILVDLKGAVQKGVYKLTASSRVRDVIELAGGLTSEADKH 114

Query: 113 SVNLAKQLQDETIVYVPAQKGEKITVVEEKANNIA-TQNSKSGKINANKADLSLQTIGS 171
 ++N A+KL DE V+YV ++GE+I+V+ + T + K++NKA L LQ I G
 30 Sbjct: 115 AINFAEKLITDEQVYVVPKQGEISVLPRSLVSGKKETASKDQSKVHINKASLEELQHIFG 174

Query: 172 VGAKRAQDILDYKRSQGGFKTIDLQVSGIGKTEKLEKRDVTD 217
 +GAKRAQDI+D RD GGPK ++DL+ VSGIGKTEKLEK+ D+ +D
 35 Sbjct: 175 IGAQRAQDIIDMRDKLGPFKALEDLRQVSGIGKTEKLKDDIFLD 220

A related GBS gene <SEQ ID 8989> and protein <SEQ ID 8990> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 5.70
 40 GvH: Signal Score (-7.5): -2.58
 Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -3.77 threshold: 0.0
 INTEGRAL Likelihood = -3.77 Transmembrane 18 - 34 (14 - 36)
 45 PERIPHERAL Likelihood = 10.40 73
 modified ALOM score: 1.25

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.2508 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

44.3/64.1% over 215aa

Streptococcus

pneumoniae
 GP|3211753| competence protein Insert characterized

60

ORF01930 (304 - 951 of 1014)
 GP|3211753|gb|AAC23741.1||AF052208(1 - 216 of 216) competence protein [Streptococcus pneumoniae]

-2452-

%Match = 25.0
 %Identity = 44.2 %Similarity = 64.1
 Matches = 96 Mismatches = 75 Conservative Sub.s = 43

```

5      90      120      150      180      210      240      270      300
   DGKGLNPLTYITRLPAITAIIVLLVLTILPSYIASFVWDPQGHK*GLHGNVLLFSK*FFWFLIKGKL*LRISKWNIIF

330      360      390      417      447      474      504      534
   MEPIVLREKTSKSHWGTGIIIVGLLLPGLLGNHFG-THHKRNLNINLEK-KVSTITTEKVPMSHVKDKVGNQIVDVX
10  ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:|
   MEALIEIKKEYKIIVICTGLGLVGGFFLLKDPAPOTPVKFTNLQAEVAASVSKDLVSEKEVKEKEEPEQDLITVDVX

15      564      594      624      654      684      714      744      774
   GAVNHFGVYSLPSQSKVTDIAIKRAGGLSNLADSKSVNLAQKIQDETIIYVAQKGEKITTVEEKANNITVCGNSKGKINL
   ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:|
10  GAVKSPGIYDLPGVGRINDAVQKAGLLEQADSKSLNIAQKVSDEALVVPVTKGEE--AVSQQTGLGTASSISKEKKVNL

15      804      834      864      894      924      954      984      1014
   NKADLSLQTSISVGAKRAQDILDYRDSQGGFKTIDDLQNVSGIGERTLEKLRQDVTDIVFSSKTYLPSIVGLPMLITS
   ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:| ||:||||:|
10  NKASLEELKQVGLGKRAQDITDREANGKFSGVDELKQVSGIGGKTIKELKDYIVTD

15      170      180      190      200      210
   
```

25 SEQ ID 8990 (GBS129) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 4; MW 43.8kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2175

30 A DNA sequence (GBS2292) was identified in *S. agalactiae* <SEQ ID 6719> which encodes the amino acid sequence <SEQ ID 6720>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -14.01 Transmembrane 215 - 231 (208 - 240)

```

35  ----- Final Results -----
      bacterial membrane --- Certainty=0.6604 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
40 
```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP: CAB12793 GB: 299109 similar to 1-acylglycerol-3-phosphate
   O-acyltransferase [Bacillus subtilis]
   Identities = 66/200 (33%), Positives = 111/200 (55%), Gaps = 10/200 (5%)
45
Query: 3 YTVLRLVIMFLIWMANGNAHYINSEDKMLKDDENYILVAPHRITFNDVYMAFAARPKQIF 62
      Y + + ++ + G Y + E+ L D +++ H + D + + P Q +
Sbjct: 2 YKFCANALKVILSRGSGVKVYNKEN--LPADSGFVIACHTSGWVDVITLGVGILPQIHY 59

50
Query: 63 NAKKELFTNRLPGWMIWMOGAPPIDREKPGQDAIRYPVVKMLKNSRSLVMPFSGSRHSKD 122
      NAKKELF N+ G ++K AFP+DRE PG +I+ P+K+LK + +FPSS+R S+D
Sbjct: 60 NAKKELFQNKWIGSFLIKIHAFPVDRENPGPSSIKTPILKLLK-EGRIVIGIFPSGSTRSSD 118

55
Query: 123 V--KGGVAVIAKMAKVRIMPAAYRGPMVFKNLLEKGERVDMNPNPDI/SDIKMDA-EGI 179
      V K G IA+M K ++PAAY+GP K L K ++ + G P+ +D + + R +
Sbjct: 119 VPLKRGAVTIAQMGKAPLIPAAAYQGPPSGKELFKKGMKMLITGEPEHQAFAHLPSKRL 178

Query: 180 A----EVSRIQEEFDRLDR 195
      A +++RI+E ++LD+
60 Sbjct: 179 AMTALNQRILEKLENDLQ 198

```

-2453-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6721> which encodes the amino acid sequence <SEQ ID 6722>. Analysis of this protein sequence reveals the following:

Possible site: 49
 5 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -11.83 Transmembrane 241 - 257 (234 - 266)
 INTEGRAL Likelihood = -4.41 Transmembrane 27 - 43 (26 - 44)
 ----- Final Results -----
 10 bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 >GP:CB12793 GB:Z99109 similar to 1-acylglycerol-3-phosphate
 O-acyltransferase [Bacillus subtilis]
 Identities = 59/198 (29%), Positives = 104/198 (51%), Gaps = 6/198 (3%)
 Query: 29 YAYLRGLVFLWVVGNAHYHHEEKMLDASENYILVAPHRTFMDPVYMAFAARPKQF 88
 Y + + + +L + G Y + E L A + + + H + D + + + P Q +
 20 Sbjct: 2 YKFCANALKVILSLRGVGVYKINKEN--LPADSGFVIACHTSGWVDVITLGVGLPQIH 59
 Query: 89 MAKKELFANRLFAWIKMCAFFIDRDKPSDPAIRYVNMILKSNRSLIMFPSSGSHSQE 148
 MAKKELF N+ + +K AFP+DR+ P P +I+ P+ +LK+ + +FPGS+R S+
 25 Sbjct: 60 MAKKELFQAKWIGSFLKKIHAFVDRNPFGPSIKTPILKLE-GEIVGIFPGSTRTSED 118
 Query: 149 V--KQGVAVIAKAKVIMFAAYCGPMSVKGALLAGEVDMTFGNPIDVSDIKRM-NDGEI 205
 V K G Ia++ K ++FAAYCGP S K L ++ + G P+ +D + + + E +
 30 Sbjct: 119 VPLKRGAVTIAQMKAPLVFAAYCGPSSGKELPKKGKMLTIGEPLAQDAFRLPKSKERL 178
 Query: 206 AEAANRIQAEFDRIDEL 223
 A + + + + +L
 35 Sbjct: 179 AAMTEALNQRIKELEHL 196

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 186/244 (76%), Positives = 212/244 (86%)
 Query: 1 MFYYTLRLTLMVFLWVVGNAHYHHEEKMLDASENYILVAPHRTFMDPVYMAFAARPKQF 60
 +FY YLR LV+FL+VW NGNAHYH+E+KML ENYILVAPHRTFMDPVYMAFAARPKQF
 40 Sbjct: 27 VFYAYLRGLVFLWVVGNAHYHHEEKMLDASENYILVAPHRTFMDPVYMAFAARPKQF 86
 Query: 61 IFMAKKELFNTNLFQGWIKMCAFFIDRDKPSDPAIRYVNMILKSNRSLIMFPSSGSHSQE 120
 IFMAKKELF NRLF NWIKMCAFFIDR+KP DAIRYV MLK SNRSL+MFPSSGSHSQE
 45 Sbjct: 87 IFMAKKELFANRLFAWIKMCAFFIDRDKPSDPAIRYVNMILKSNRSLIMFPSSGSHSQE 146
 Query: 121 KDVGKGVAVIAKAKVIMFAAYCGPMSVKGALLAGEVDMTFGNPIDVSDIKRMNDGEIA 180
 ++VKGKGVAVIAK+AKV+IMFAAY+GPM K LL G RVIDM FGNPIDVSDIKRM+ EGIA
 50 Sbjct: 147 QEVKGVAVIAKAKVIMFAAYCGPMSVKGALLAGEVDMTFGNPIDVSDIKRMNDGEIA 206
 Query: 181 EVSRRIQEEDFDRIENSTYDGGKINPLTYITRLPLAIIAIVLLMLTIFSLYASFWMD 240
 EV+ RIQ EFD+D E + GK NPLTY+YRLPL ++ +V+L+LT++FSY+ASFWN+
 55 Sbjct: 207 EVANRIQAEFDRIIDELAPFQPKARNPLTYITRLPLGLVILVLLMLTIFSLYASFWN 266
 Query: 241 PQKH 244
 P KH
 Sbjct: 267 PDKH 270

60 SEQ ID 6720 (GBS171) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 36 (lane 2; MW 25kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 3; MW 49.8kDa).

-2454-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2176

A DNA sequence (GBSx2293) was identified in *S. agalactiae* <SEQ ID 6723> which encodes the amino acid sequence <SEQ ID 6724>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3268 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CA11810 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
Identities = 113/244 (46%), Positives = 173/244 (70%), Gaps = 2/244 (0%)

Query: 6 LKENERIDQLPSTDVVKIQNKVEFVSYSIDSVLLSRFPKLP-SRGLIVDLCSGNAGVLF 64
L++ER+D L+ +D+KIIQ+ VF++S+D+VLLS+F +F +G IVDLCSGNAGV L L
Sbjct: 4 LHDDERLDYLAEQMKIIQSPVTFAPSLDVLVLSKFAVYPIQKGVLDLCTNGSVPLLL 63

Query: 65 STKNTATIIETELQESLADMAKESIKANKLEKQVTIMNDLQKLLDHVQSRNVLMLCNF 124
ST++ A I+ +E+QR L DMA RS++ NKL+ Q+ +I+DLKN+ + + + D++ CNP
Sbjct: 64 STSKADILGVEIQERLADMAVRSVEYNKLDQIQIHDLLQKMFELGHNRVYDVVTYCNP 123

Query: 125 PPFKASSTSKKQLSPHYLLARHEITTNLRICQIAQHALAKTKGRILAMVHRFDRLFLEITD 184
PYFK + +++N+ + H +ARHEI L ++ ++ LK G+ A+VHRP R LRI +
Sbjct: 124 PPFKTPKQTEQNMENHLRIARHEIHCTLEDVIVSSKLLKQGGKALVHRPGRLLRIFEL 183

Query: 185 MRQFNLAPEKRIQFVYPKLGKDANMLLEAKLDGSGTEGMKILPELVHQNDGYTIFDI 244
M+ + + PER-QEVYPK GK+AN +L+E IK G + +KILPEL V+ + +YT+ I I
Sbjct: 184 MKAYQIEPKRVQFVYPKQKEANTILVEISIKGGRFD-LKILPELVFYDEQNYTKIRITI 242

Query: 245 YFGE 248
+G+
Sbjct: 243 LYGD 246
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6725> which encodes the amino acid sequence <SEQ ID 6726>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2183 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 200/257 (77%), Positives = 228/257 (87%), Gaps = 3/257 (1%)

Query: 1 MIDTILKENERIDQLPSTDVVKIQNKVEFVSYSIDSVLLSRFPKLP-SRGLIVDLCSGNAGV 60
MI ILKE ERIDQLPS+DV IIQNK+VFSYSIDSVLLSRFPK+PS+GLIVDLCSGNAGV
Sbjct: 1 MIKAILKEGERIDQLPSSDVGIQNKQVFSYSIDSVLLSRFPKMPSEGLIVDLCSGNAGV 60

Query: 61 GLFASTKNTATIIETELQESLADMAKESIKANKLEKQVTIMNDLQKLLDHVQSRNVLMLCNF 120
GLFAST+T A I+E+RLQ LADMA +RSI+L+N+LE QVTIMI DDLKRLN+HV RS VDLN
Sbjct: 61 GLFASTKTKAIVEVRLQERLADMGQRSIQLNQLREQVTIMICDLKRLN+HVRSQVDLM 120

Query: 121 LCNFPFKASSTSKKQLSPHYLLARHEITTNLRICQIAQHALAKTKGRILAMVHRFDRLFLE 180
```


-2455-

LCNPPYFK+ E+SKN+S HYLLARHE+TTNL EICQ+A+HALK+ GR+AMVHRPDRFLE
 Sbjct: 121 LCNPPYFKSHSSKKNVSHYLLARHEVTTNLEICQVARHALKSNGLAMVHRPDRFLE 180

Query: 161 IIDTMRQFNLAQRIQFVYKLGKIDANMLLIRAIKDGSGTGMKILPFIWVHQDNGDYTET 240
 IID+R LAPIR+QFVYKLGK ANMLLIRAIKDGSGTGMKILPFIWVH++NG+YI+
 Sbjct: 161 IIDSRLANGLAQRIQFVYKLGKSNMLLIRAIKDGSGTGMKILPFIWVHKENGSDYTH 240

Query: 241 IFDIYFGENGK---SHD 254
 IF+IYFG K +HD
 Sbjct: 241 IFETIYFGAASGKGNHD 257

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2177

- 15 A DNA sequence (GBSx2294) was identified in *S. agalactiae* <SEQ ID 6727> which encodes the amino acid sequence <SEQ ID 6728>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1512 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11811 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 40/82 (48%), Positives = 63/82 (76%)

Query: 7 YMYLVLECSGDTLYTGYTIDVKRRILNTHNGKAKYTRARLPVKLLYSEAFNSKQEAQRAE 66
 + YV+ C D + Y GYT D+ +R+ THN GKAKYT+ R FV+L+++E+F++K+EAM+AE
 Sbjct: 7 FFFVVKCKDINSYAGYTNDLHKRVKTHNDGKAKYTKVRPVELTFABSFSTKREMAQAE 66

Query: 67 ALFFKQTRQAKLTYIKQHKNQ 88
 FK+ TR+ K YI++ +N +
 Sbjct: 67 YYPFKLTRKKKELYIEKRNK 88

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6729> which encodes the amino acid sequence <SEQ ID 6730>. Analysis of this protein sequence reveals the following:

40 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1838 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/84 (71%), Positives = 67/84 (79%), Gaps = 1/84 (1%)

50 Query: 6 AYMYVLECSGDTLYTGYTIDVKRRILNTHNGKAKYTRARLPVKLLYSEAFNSKQEAQRAE 65
 AYMYVLECD TLYTGYTTD+K+RL THN GKAKYTR RLFP LLY E F+SK+ AM A
 Sbjct: 6 AYMYVLECDVKTLYTGYTTDLKKRLATNAGKAKYTRYRLPVSLLAYTEVDFDSKGAAMGA 65

Query: 66 EALF-EKQTRQAKLTYIKQHKNQ 88
 EALF K+ETR KL YI H+ E+
 Sbjct: 66 EALFKKRKRTSKLAYIATHOKEK 89

-2456-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2178

A DNA sequence (GBSx2295) was identified in *S. agalactiae* <SEQ ID 6731> which encodes the amino acid sequence <SEQ ID 6732>. This protein is predicted to be autoaggregation-mediated protein (dead). Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2287 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD20136 GB:AF09:502 autoaggregation-mediated protein
[Lactobacillus reuteri]
Identities = 289/504 (57%), Positives = 366/504 (72%), Gaps = 18/504 (3%)

Query: 1 MKFTENLNSQDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAGFLPTL 60
MKF+EL LS +L A++++G+ E +PIQE TIP+ LEGKDVIGQAQTGTGKTAAGFLP +
Sbjct: 1 MKFSELGLSDLLKAIKRSGYEEATPIQBQTI PMVLEGKDVIGQAQTGTGKTAAGFLPTI 60

Query: 61 NKIHTEINDTQALIIAPITRELAVQSQBELFRFGREDKGVKRVSVYGGSSGIEKIKALRSQA 120
+ TE+ IQA+II+PIRELA+Q+QREL+R G+DK V+V+ VYGG+ I +QIK+L+
Sbjct: 61 ENVDTENPNTQALIIISPTRELAITQDELYRLGDKKHVRVQVYVGGADIRRQIKSLKQHP 120

Query: 121 HVVVGTGPRGLLDLKKRALKLNHITLILDEADEMLAWGFLIEDIEAIIISRVPTGRTLLF 180
++VGTGPRL D I R +KL+HI+TL+LDEADEMLAWGFLIEDIE+II P+ RQTLF
Sbjct: 121 QILVGTGPRLRDHINRHTVKLHITLVLDEADEMLAWGFLIEDIESIKETPDORGTLLF 180

Query: 181 SATMPDPPIKRIQVFKMKDPEHVKIKATELTNNVVDQYTVRVKENKFDIMTRLAMDVDQPE 240
SATMP IKRIGV+FM DPE V+IKA ELT VDQYTVR ++ EKFD MTRL+DV P+
Sbjct: 181 SATMPPIIKRIQVQFMSDPETVRIKAKELTTLVDQYTVRVARDYKFDIMTRLIDVQDPD 240

Query: 241 LSIIVPGRTIKRRVDELTRGLKLRGFRABGIGDLDQNKRLRVIRDPIQNDHIDILVAIDVAA 300
L+IVPGRTIKRRVDEL++GL RG+ A GIGDL Q+KR +++ FIM+ +DILVAIDVAA
Sbjct: 241 LTIIVPGRTIKRRVDELSKGLIARGVNAIGHGLTDQIKRSKIMMKFQNNELDILVAIDVAA 300

Query: 301 RGLDISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSQSITFVS PNEMGYLTIIENLTKK 360
RGLDISGVTHVYNYDIP DP+SYVHRIGRTGRAG G S+TFV+PNEM YL IE LT+
Sbjct: 301 RGLDISGVTHVYNYDIPSDPDSYVHRIGRTGRAGHGVSLTFVTFNEMGYLHETKLRV 360

Query: 361 RMTGKMPATASEFOAKKKVALKRIARDFED-QLRVSK--FDKFKADALSLATQYTFEEL 417
RM +KF TA EAF+ ++A F D EL+++ D+++ A+L + +L
Sbjct: 361 RMLPLKPTAEAEAFKQ-----QVAFPNIDIELLAQSDTRYEEAARLLETHNATDL 413

Query: 418 ALVYLSLITVDQPESELPEVEITREKLPFFKPGGGGKGGKGGNGRGGD--RRNRDGRDR 475
+L+ ++ S V+IT E+PLP + G R N GG+ RR+N R +
Sbjct: 414 VALLNNMTKEAASESVFKITPERLPLPRNRGRN--NRNGNRNNSHGNNHYRRKNFRFRQH 471

Query: 476 GNRDRDRDQ----SRCDPKRRDDK 495
G+ D+ G SR F R K
Sbjct: 472 GSIRNNHGKSHSGRRSHFNIRHK 495

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6733> which encodes the amino acid sequence <SEQ ID 6734>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

-2457-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1108(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 430/545 (78%), Positives = 463/545 (84%), Gaps = 24/545 (4%)

Query: 1 MKFTLNLSQDILSAVEKAGFVSPSPQIEMTIPALBGRDVGQAQTGTGKTAAFGPLTL 60
 +KFTS NLSQDI SAV AGF + SPQIEMTIPALBGRDVGQAQTGTGKTAAFGPLTL
 Sbjct: 1 LKFTFNLSQDIQSAVVTAGFEKASPIQEMTIPALBGRDVGQAQTGTGKTAAFGPLTL 60

Query: 61 NKIHTEDNTIQAIIAPTR ELAVQSQEELFPFRG+KGVKVRSVYGOSSIERQIKALRSGA 120
 NKI T +N IQAIIAPTR ELAVQSQEELFPFRG+KGVKVRSVYGOSSIERQIKAL+SGA
 Sbjct: 61 NKIRITNENIIQAIIAPTR ELAVQSQEELFPFRG+KGVKVRSVYGOSSIERQIKALKSGA 120

Query: 121 HVVVGTGRLLDLIKRKALKLNHIETLILDEADENLNMGFLEDEIATISRPVETRQTLLF 180
 H+VVGTPGRLLDLIKRKAL L+H+ETLILDEADENLNMGFLEDEIATISRPV RQTLLF
 Sbjct: 121 HIVVGTGRLLDLIKRKALILDHVETLILDEADENLNMGFLEDEIATISRPVADRQTLLF 180

Query: 181 SATMPDPKIRIGVKFMKDPHEVVKIKATELTNNVVDQITVRVKEKNEFDTMIRLMDVQPE 240
 SATMP PIK+IGVKFMKDPHEV+IK ELTNVVDQITVRVKE EKFDPMIRLMDV+QPE
 Sbjct: 181 SATMPAPIKQIGVKFMKDPHEVQIKNELTNNVVDQITVRVKEQKFDPMIRLMDVQPE 240

Query: 241 LSVIFGRTIRRVDELTIRGLKRGFRABGIHGDLDQNKRLVIRDFKNDHIDILVATDVAA 300
 LSVIFGRTIRRVDE+TRGLKRGFRABGIHGDLDQNKRLVIRDFKND IDILVATDVAA
 Sbjct: 241 LSVIFGRTIRRVDELTIRGLKRGFRABGIHGDLDQNKRLVIRDFKNDQIDILVATDVAA 300

Query: 301 RGLDISGVTHVINYDI PQDPESIVHRIKRGORAGKSGSITVSPNEMGYLTIENLTKK 360
 RGLDISGVTHVINYDI QDPESIVHRIKRGORAGKSG+SITVSPNEMGYL+TIENLTKK
 Sbjct: 301 RGLDISGVTHVINYDI PQDPESIVHRIKRGORAGKSGSITVSPNEMGYLSMIENLTKK 360

Query: 361 RMTGKIPATASAPQAKKKVALKRIARDPFDQELVSKFOKFKADALELATQYTPPEALY 420
 +M +PATA SAPQAKKKVALK+I RDF Da +S FDKFK DA++LA ++TPPEALY
 Sbjct: 361 QMKFLAPATASAPQAKKKVALKRIERDPADETIRSNFDFKFGDAVQLAAETPEALY 420

Query: 421 VLSITVQDPESLPEVEITREKPLPFKPSGGGF---KKGGRG--NCRGDRRRNRDGRD- 474
 +LSITVQDP+SLPEVEI REKPLPFK GGG GKGRG N GDRR RGRD
 Sbjct: 421 ILSITVQDPDSLPEVEIAREKPLPFKIVGGGHGNGKGGGRGDRNRGDRRGYVGRDNR 480

Query: 475 -----RNRDRDRGSRCDKFRDDEKPKNKRQENKIPHNTSSEKQTGFVI 522
 R RD D DEK+ + KD +E K SS K TGFVI
 Sbjct: 481 RDRDGRRRRQGRDGRDGHGSGNRDFKRSKRSKDFFNKEKK-----SSAKNTGFVI 534

Query: 523 RNKG 527
 R+KG+
 Sbjct: 535 RHKE 539

A related GBS gene <SEQ ID 8991> and protein <SEQ ID 8992> were also identified. Analysis of this protein sequence reveals the following:

RGD motif 471-473

The protein has homology with the following sequences in the databases:

56.9/74.7% over 494aa

Lactobacillus reuteri

GP|4409804| autoaggregation-mediating protein Insert characterized

CRP01926(301 - 1785 of 2184)

GP|4409804[gb]|AAD20136.1|[AF091502(1 - 495 of 497) autoaggregation-mediating protein
 [Lactobacillus reuteri]

%Match = 37.3

%Identity = 58.8 %Similarity = 74.6

Matches = 290 Mismatches = 118 Conservative Sub.s = 78

42 72 102 132 162 192 222 252

-2458-

```

IRHYITKEIPSEAAVAF*IDKL*TLILYRNWVFIAFPLPSEATNRTSRL*KRVIY*IDLILYLPFRVCVTLRLSEKITN
282      312      342      372      402      432      462      492
KGS*GSFALFSRKEKHLKPTRELNLQDILSAVEKAGFVEPSIQEWMTPLALRGKQVIGQAQTGKTAAQPLPTLNKIH
5      :||:| | | | | :| :|:::| :| :||| | | :| | | | | | | | | | | | | | | :| :|
      MKPSKLGSLSLKAIKIKSGYBEATPIQKQTIPIVWLGKQVIGQAQTGKTAAQPLPTLRVD
      10      20      30      40      50      60

522      552      582      612      642      672      702      732
TEDTITQALIIATPTELAVQSGELPLFRGRDGVKVRVSYGSGSIEKQIKALRSGAHVVVGTGPRILLDIATKRLKALIMHI
10      ||: ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      TENPNIGAIITISPTRELAICTQSELYTLKGDKHVRVQVYGGADIRKQIKSLKQHPQILVGTGRLADHINKRETVLKHDI
      80      90      100      110      120      130      140

152      182      222      262      302      342      382      422
ETILLDEADEMLNGPLEDIEALISFVPETRQTLFSATMPDPPIKRGVQPMKDPESHVKIKATELTVWVVDVYVVRKEN
15      :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      KTVLDEADEMLNGPLEDIESIIKETFDGRQTLFSATMPPEIKRGVQPMKDPESHVKIKATELTVWVVDVYVVRKEN
      160      170      180      190      200      210      220

202      242      282      322      362      402      442      482
EKFTPTWRLMDVQPELSIVFGRTRGRVDELTRGLKRGFRAEGIHGDLQNKRLRVIRDFHNDIIDLIVATDVAARGLD
20      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      EKFDPTWRLMDVQPELSIVFGRTRGRVDELKGLIARGTNAAGIHGDLTQDKRSKIMKPFQNELDILVATDVAARGLD
      240      250      260      270      280      290      300

252      292      332      372      412      452      492      532
ISGVTYVNDIPQDPESYVHRIGRTGRAGKSGQSITFVSPNMGYLTIIENLTKKRMTGMKFAASEAPCAKKKVALKR
25      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      ISGVTYVNDIPQDPESYVHRIGRTGRAGHHGSLTFTVTPNEMDYLHEIEKLTRVRLPLKPFPTAEAF--KGQVA---
      320      330      340      350      360      370

302      342      382      422      462      502      542      582
IARFDN-QBLVSK--FDKFKADALELATQYTPRELALYVLGLTVQDPESLFEVEITREKPLFPKPSGGQSPKGGKQGRNG
30      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      --SAPNDIDELAQDSDTRYKEAAEKILETHNATILVAALINMTKEAASEVPVKITPRLPRNRKRNKRN--RNS
      390      400      410      420      430      440      450

352      392      432      472      512      552      592      632
RGGD--RRNRDGDGRGNRRDRDGR---SRCDPKREDKPKKCNRRGNKXPHKNTSSEKQTFVIRNKGD*EDYEKG
35      ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      HGGNHYRKKKFRRHQHGSHRNDNHGKSHSSRHSFNIRHRKEN
      470      480      490

```

45 There is also homology to SEQ ID 4454.

SEQ ID 8992 (GBS307) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 7; MW 62kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 2; MW 86.7kDa).

45 The GBS307-GST fusion product was purified (Figure 208, lane 9; Figure 225, lane 10-11) and used to immunise mice. The resulting antiserum was used for FACS (Figure 272), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2179

55 A DNA sequence (GBSx2296) was identified in *S. agalactiae* <SEQ ID 6735> which encodes the amino acid sequence <SEQ ID 6736>. This protein is predicted to be outer membrane protein (yaeC). Analysis of this protein sequence reveals the following:

-2459-

Possible site: 19
>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CB73036 GB:ALL39076 putative periplasmic protein [Campylobacter
jejunii]

Identities = 89/237 (37%), Positives = 132/237 (55%), Gaps = 3/237 (1%)

Query: 40 ITVATYSKPTSTFLDLVKDNVKGKGYTLKVMVSDYIQANLAIENKHEHDANLLQHEFFMS 99
IT+ P + L+L+KD+ K KGY LK+V SDYI N ALE KE DANL QH+ F+

Sbjct: 23 ITIGATPNPFGSLLELMKDDFKNGYELKIVFSDYILNFRALBEKEDANLYQHKFPLE 82

Query: 100 IFNKENDGHLVSIPTIYHSLAGFYQHLLNTAELKDGAIVAPSDPANMTRALLLQK 159
+N + +L++ TP+ + G Y + +KN+ LK+GR+VAIP+D N +RAL LL++ K

Sbjct: 83 EYNLCKGSLNIATTPVLIAFVGVTYSRKINLENLKEGARVAIPNDATNESRALELEKAK 142

Query: 160 LITLQNTSKKTKAIEDIITNPKKLRIEFPVALLNLNQAYFEDLVFNFPQYVTKINLVPR 219
LI L + KT DI NPCKL+ + L +A + D+ + L P +

Sbjct: 143 LIEKNQNTLKTFL--DINKPKLKFIEFLKAAQLFRALDDVIALINSFALGAGLNFSK 200

Query: 220 DRLLYKKKPDIFRAGALVAREDNQNSDKIKVLKEVLTSKEIRHYITKIPSEAAVAF 276
D + E K + + +V R + KNS+K KV+ E+L S + + I + AF

Sbjct: 201 DTIFREDK-NSPYVWTVVBSGKNSSEKTKVIDEILSDSKFALINEHYKDLILIPAF 256

SEQ ID 6736 (GBS126) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 7; MW 32kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2180

A DNA sequence (GBSx2297) was identified in *S. agalactiae* <SEQ ID 6737> which encodes the amino acid sequence <SEQ ID 6738>. This protein is predicted to be probable permease of ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.99	Transmembrane	190 - 206 (187 - 215)
INTEGRAL	Likelihood = -8.44	Transmembrane	25 - 41 (16 - 45)
INTEGRAL	Likelihood = -6.48	Transmembrane	69 - 85 (68 - 90)
INTEGRAL	Likelihood = -3.77	Transmembrane	90 - 106 (88 - 109)
INTEGRAL	Likelihood = -1.44	Transmembrane	145 - 161 (145 - 161)

----- Final Results -----

bacterial membrane --- Certainty=0.5798 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG08869 GB:AR004963 probable permease of ABC transporter
[Pseudomonas aeruginosa]

Identities = 80/206 (38%), Positives = 127/206 (60%), Gaps = 4/206 (1%)

Query: 15 SFWEINMLGLTLLCLFIAPFTGILLFSLRKYLIKHSLAYQLNLFGLTRSRVFPFLIF 74
+FW MIG +L+ + P G+LLF + + Y LL+L + LR3+EF+I

Sbjct: 24 TFW----MLGSLIFTVVLGLFLGVLLFLTGPRQPFQKAVYTLLSUNVNILRSLEFIFL 79

-2460-

Query: 75 IFILLPLNRLIPGTSFGTIAAILPLTLVSLSLYARYVEQALNITPQVVVDRAISLGANKR 134
 ++IPL LI GTS G AI PL ++ +AR V3 AL ++ + + +GA+ R
 Sbjct: 80 LIVMPLTLVLTGTSIGVAGAIPLLVGATPFFARLVTALREVVDKGIIEATQAMGASTR 139

Query: 135 QIIYFLIPSIIKIDLVLSFTATAISILGYSTINGVIGAGLGSYAYRPGYQRYDYPVNYL 194
 QII+ L+P + ++ + T TAI+++ Y+ + GV+GAGLG+ A RGYQ + VM +
 Sbjct: 140 QIINALLPEARPGIIAAITVTAITLVSFYTAGVVGAGLGLDLAIRPGYQRPCTDVMV 199

Query: 195 IVVLFIIYVFLQLSGIFYFIANYRSK 220
 VV+ +I V ILQ++G + +SRK
 Sbjct: 200 TVVMILLVQLITVQDKLVVHFSRK 225

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2181

A DNA sequence (GBSx2298) was identified in *S. agalactiae* <SEQ ID 6739> which encodes the amino acid sequence <SEQ ID 6740>. This protein is predicted to be ABC transporter, ATP-binding protein (oppF). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.5454 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9333> which encodes amino acid sequence <SEQ ID 9334> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22280 GB:U32744 ABC transporter, ATP-binding protein
 [Haemophilus influenzae Rd]
 Identities = 62/174 (35%), Positives = 104/174 (59%), Gaps = 2/174 (1%)

Query: 1 MKMINGLIPIYDKNIYQGEKVSFSDNKLRCMRKDIAYIFQNNHLLAGSVVYHLALVY 60
 ++ +N L G++ G E+ SD +L R+ I IFQ+ NLL+ +V+ ++AL
 Sbjct: 48 IRCVNLEKPTSGSVIVDGVLTSLGORELVLARQIGMIPQHFNLSSRTVFENVALPL 107

Query: 61 KLEHOKVN--HDAINDILDFGLMDLKQVKCHSLSGGQQQKVAIAMAVALQKPKLILCDEI 118
 +L + + I +LD +GL ++ +LGGQ+Q+VAIA A+ PK++LCDE
 Sbjct: 108 ELSEFSKAKIQKRTITALLDLVGLSEKRDYPSNLGGGQKQVVAIAAALASDPKVLICDEA 167

Query: 119 SSALDTNSKEIFNLLSDLEKYGISILIMAHLLSHLKOYCDRVMLDHQCTIVD 172
 +SALD ++ I LL ++ GI+IL+I H + +KQ CD+V ++D +V+
 Sbjct: 168 TSALDPATVQSLKLKLEINRTLGITILLITHREMEVVKQICDQVAVIDQGRIVE 221

There is also homology to SEQ ID 76.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2182

A DNA sequence (GBSx2299) was identified in *S. agalactiae* <SEQ ID 6741> which encodes the amino acid sequence <SEQ ID 6742>. Analysis of this protein sequence reveals the following:

Possible site: 21

-2461-

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2183

A DNA sequence (GBSx2300) was identified in *S.agalactiae* <SEQ ID 6743> which encodes the amino acid sequence <SEQ ID 6744>. Analysis of this protein sequence reveals the following:

15 Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.0904 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9741> which encodes amino acid sequence <SEQ ID 9742> was also identified.

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AA07515 GB:AF034138 unknown [Bacillus subtilis]
 Identities = 74/125 (59%), Positives = 92/125 (73%)

30 Query: 5 MGIFSGIAGNASQMDTKVENQLSDILISDFQVQLAVTILRLDILVFTNYRLILVDKQGV 64
 MG GL+GNAS + T V+ +L+ IL+ E+V+ A+ L+RLIVFT+ RLILVDKQG+T
 Sbjct: 1 MGFIDGLGNASTLSTAAVQELAHILLBGEKVEAFAFKLVRDLIVFTDKRLILVDKQGIT 60

Query: 65 GKKVSYNSIPVASISRFTVETSGHFDLDAELKIWISSAIEPAEVLQFNDRNIVSIQKAL 124
 GKK + SIPY SISRF+VET+G FDLD+ELKIWI S A PA QPK D +I IQK L
 Sbjct: 61 GKKTFFQSIPIKYSIRSPSVETAGRFDLSELKIWISSGAELPAVSKQPKKDESIYDIQKVL 120

Query: 125 ATAVL 129
 A +
 Sbjct: 121 AAVCM 125

40

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2184

- 45 A DNA sequence (GBSx2301) was identified in *S.agalactiae* <SEQ ID 6745> which encodes the amino acid sequence <SEQ ID 6746>. Analysis of this protein sequence reveals the following:

 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.0921 (Affirmative) < succ>

-2462-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9331> which encodes amino acid sequence <SEQ ID 9332> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA74739 GB:Y14370 peptide chain release factor 3
 [Staphylococcus aureus]
 Identities = 274/462 (59%), Positives = 349/462 (75%), Gaps = 9/462 (1%)

Query: 1 MDIERQGISVTSSVMQFDYAGKRVNLTDPGHEDFSEDYTRILMAVDAVMVDSAGKI 60
 M +E+RGISVTSSVMQFDY +NLTDPGHEDFSEDYTRILMAVD+AVMV+D AKG+
 Sbjet: 57 MKVEQERGISVTSSVMQFDYDYEINLTDPGHEDFSEDYTRILMAVDSAVMVDCAGV 116

Query: 61 EAQTKKLFEVVKHNRNIPVFTFINKLDRDGREFLLLEELSEVLGIASYPMNPMIGMKSF 120
 E T KLF+V K R IP+FTFINKLOR G+EP +LL+E+SE L I +YPMNPMIGMK+SF
 Sbjet: 117 EPPTLKLKFKVCKMRGIPFTFINKLORVGKEFFPELLDEISLNTIETPMNPMIGMKQSF 176

Query: 121 EGLYDLNKRLELYKGRFASIEDG----DQLFANNPFYECVKREDIELLQAGNDFSE 175
 G+ D +K +E ++ +E + D D N+ +EQ E++ L++EAG F
 Sbjet: 177 FGIIDRKSRTIEPRDEENILHLANDOFEECHAITNDSDFEQAIRLEMLVEEAGEAFDN 236

Query: 176 QAILDGLDTPVFFGSALTNFGVQTFDITFLEFAPEPHGHKTTGNGVIDLAKDFSFGVFK 235
 A+L GDLTPVFFGSAL NFGVQ FL+ ++EAP P+ +T E + P FSGF+FK
 Sbjet: 237 DALLSGDLPVFFGSALANFGVQNFNLAIVDFAPMPNARQTKENVBSPFDDSGFGIFPK 296

Query: 236 IQANNDPRHRDRIAPRVISGEFERGMVNLTRTKGAKLGNVQFMAES-RENVNNAVA 294
 IQANNDP+HRDRIAP+R+VSG FER + + L +K S+V + + + + + V +AVA
 Sbjet: 297 IQANNDPRHRDRIAPRMVVSAGAPER-VAMLLCNVLKSRSHVQRHLQITIKIVNNAVA 355

Query: 295 GDIIQVYDGTGYQVQDITLVGNKFFPEFLPTFPELFMKVSARKVMQKSFHKGIEQLV 354
 GDIIQ+YDGT YQ+GDITL GK + F+ LP FIPE+FMKVSARKVMQK FHRGIEQLV
 Sbjet: 356 GDIIQLYDGTGYQVQDITLVGGKQTSYFQOLQPTFPELFMKVSARKVMQK FHRGIEQLV 415

Query: 355 QBSAQLYNYQYTGEMVLGAVGQQLQFEVFKHRMBEGYNAEVVMTPMKKTKVRN--INSD 412
 QBSAQL YK T + +LGAVGQQLQFVFP+HRM+ EYN +VM P+G+K RW N D
 Sbjet: 416 QBSAQLYNYKTLHTNQLILGAVGQQLQFVFEHRMKNEYNVDMVEPVRKRIARWDIENEDQ 475

Query: 413 LDERMSSSRNLAKDRFDQFVFLFENDPALRWFAKYPPVVL 454
 ++M++SR+IL KDR+D VFLFN+FA RWF +K+P++IL
 Sbjet: 476 ITDKNTSRSLVVKRYDGLVFLFMEFATRWFEKFFPEIL 517

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6747> which encodes the amino acid sequence <SEQ ID 6748>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2070 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 447/458 (97%), Positives = 455/458 (98%)

Query: 1 MDIERQGISVTSSVMQFDYAGKRVNLTDPGHEDFSEDYTRILMAVDAVMVDSAGKI 60
 MDIERQGISVTSSVMQFDYAGKRVNLTDPGHEDFSEDYTRILMAVDAVMVDSAGKI
 Sbjet: 57 MDIERQGISVTSSVMQFDYAGKRVNLTDPGHEDFSEDYTRILMAVDAVMVDSAGKI 116

Query: 61 EAQTKKLFEVVKHNRNIPVFTFINKLDRDGREFLLLEELSEVLGIASYPMNPMIGMKSF 120
 EAQTKKLFEVVKHNRNIPVFTFINKLDRDGREFLLLEELSEVLGIASYPMNPMIGMK+P
 Sbjet: 117 EAQTKKLFEVVKHNRNIPVFTFINKLDRDGREFLLLEELSEVLGIASYPMNPMIGMGRAF 176

-2463-

Query: 121 EGLYDLINKRLELYKGDERFASITRDGDLFANNPFYEQVKEDIELLQEAQNDPSEQAILD 180
 EGLYDLINKRLELYKGDERFASITRDGDLFANNPFYEQVKEDIELLQEAQNDPSEQAILD
 5 Subjct: 177 EGLYDLINKRLELYKGDERFASITRDGDLFANNPFYEQVKEDIELLQEAQNDPSEQAILD 236

Query: 181 GDLTPVFFGSALTINFGVQTFLLDTFLFPAPEPHGHKTTTEGNVDDPLAKDFSGFVFKIQANN 240
 GDLTPVFFGSALTINFGVQTFLLDTFLFPAPEPHGHKTTTEGNV+DPLAKDFSGFVFKIQANN
 10 Subjct: 237 GDLTPVFFGSALTINFGVQTFLLDTFLFPAPEPHGHKTTTEGNVDDPLAKDFSGFVFKIQANN 296

Query: 241 DPHRDRIFAIVRIVSGEPERGMGVNLTITCKGAKLSNVTFQMASSREBNVNAVAGDIIGV 300
 DP+HRDRIFAIVRIVSGEPERGMGVNLTITCKGAKLSNVTFQMASSREBNVNAVAGDIIGV
 10 Subjct: 297 DPHRDRIFAIVRIVSGEPERGMGVNLTITCKGAKLSNVTFQMASSREBNVNAVAGDIIGV 356

Query: 301 YDTGTIYQGDITLTGKDKFEPFELPTFTPELPMKVSAKIVMKQKSFHGKIEQLVQBSAIQ 360
 YDTGTIYQGDITLTGKDKFEPFELPTFTPE+PMKVSAKIVMKQKSFHGKIEQLVQBSAIQ
 15 Subjct: 357 YDTGTIYQGDITLTGKDKFEPFELPTFTPELPMKVSAKIVMKQKSFHGKIEQLVQBSAIQ 416

Query: 361 LYNYQTGEIMLGAVGQLQFEVYKIRMSGEYNABVNTPMKKITVRWINSDDLERMSSS 420
 LYNYQTGEIMLGAVGQLQFEVYKIRMSGEYNABVNTPMKKITVRWIT+ DDL+RMSSS
 20 Subjct: 417 LYNYQTGEIMLGAVGQLQFEVYKIRMSGEYNABVNTPMKKITVRWINSDDLERMSSS 476

Query: 421 RNILAKDRFDQPVFLFENDFALRWADKYDVKLEEM 458
 RNILAKDRFDQPVFLFENDFALRWADKYDVKLEEM
 25 Subjct: 477 RNILAKDRFDQPVFLFENDFALRWADKYDVKLEEM 514

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2185

A DNA sequence (GBSx2302) was identified in *S. agalactiae* <SEQ ID 6749> which encodes the amino acid sequence <SEQ ID 6750>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.3061 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP: AAC38046 GB: AF000954 No definition line found [Streptococcus mutans]
 Identities = 122/142 (85%), Positives = 138/142 (96%)

Query: 1 MLEFAAQKTGKDKEMAVTFVTVNSRSHLNLEYRDTPTDVISLEYKPEVDISFDEDL 60
 +LEFAAQKTGKDKEMAVTFVTVNSRSHLNLEYRDTPTDVISLEYKPE +SFDEDL
 45 Subjct: 23 ILEFAAQKTGKDKEMAVTFVTVNSRSHLNLEYRDTPTDVISLEYKPEVDISFDEDL 82

Query: 61 AENPELAEMLEDVDSYIGELFISIDKAKSQAKEYGHSYERMGFLAVRGFLHNGYDHYT 120
 A++P+LAH+L +FD+YIGELFIS+DKA+BQA+KYGHS+ERMGFLAVRGFLHNGYDHYT
 50 Subjct: 83 ADDPDALVLTETFDATYIGELFISVDKAREQAKEYGHSYERMGFLAVRGFLHNGYDHYT 142

Query: 121 PEREKEMFSLQEEILAYGLKR 142
 P+SEKEMFSLQEEILAYGLKR
 Subjct: 143 PQSEKEMFSLQEEILAYGLKR 144

55 There is also homology to SEQ ID 120.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2464-

Example 2186

A DNA sequence (GBSx2303) was identified in *S.agalactiae* <SEQ ID 6751> which encodes the amino acid sequence <SEQ ID 6752>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -15.39    Transmembrane    108 - 124 ( 100 - 131)
    INTEGRAL    Likelihood = -8.92     Transmembrane    61 - 77 ( 52 - 82)
    INTEGRAL    Likelihood = -5.36     Transmembrane    41 - 57 ( 40 - 60)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.7156(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  The protein has homology with the following sequences in the GENPEPT database.
      >GP:AA38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]
      Identities = 107/133 (80%), Positives = 121/133 (90%), Gaps = 2/133 (1%)

Query: 1 MDLNDN--NHKKWNRITLSSMEFAVTGIFTAFKERNRKHLVSAILVLGLTQVSM 58
      MDL DN + KWKWNRITLSS+EFA+TGIFTAFKERN+KH VSA+L ++AGL P+VS+
20  Sbjct: 3 MDLRNKSQKWKWNRITLSSLEFALTGIFTAFKERNMKHVASALLAVIAGLVFKVSV 62

Query: 59 IEWLFLLLSIFLVITFEIVNSAIEINVVDLASDYHFSMLAKNAKMAAGAVLVVSLFAVLV 118
      +EWLFLLLSIFLVITFEIVNSAIEINVVDLAS+YHFSMLAKNAKMAAGAVLV+SA L
25  Sbjct: 63 IEWLFLLLSIFLVITFEIVNSAIEINVVDLASDYHFSMLAKNAKMAAGAVLVISGFAALT 122

Query: 119 GLIIFPKILALL 131
      GLIIF+PKI LL
30  Sbjct: 123 GLIIFVFKIWFL 135
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6753> which encodes the amino acid sequence <SEQ ID 6754>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -10.67    Transmembrane    63 - 79 ( 41 - 84)
    INTEGRAL    Likelihood = -7.32     Transmembrane    110 - 126 ( 105 - 129)
    INTEGRAL    Likelihood = -5.41     Transmembrane    43 - 59 ( 41 - 62)

40  ----- Final Results -----
      bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the databases:

```

45  >GP:AA38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]
      Identities = 104/135 (77%), Positives = 119/135 (88%)

Query: 1 MALHNDNNTTKRWKNRITLSSLEFALTGVFTAFKERNLRSHLLSACLACVAGLFFSISA 60
      M L DN +++KWKWNRITLSSLEFALTGVFTAFKERN++ H +SA LA +AGL F +S
50  Sbjct: 3 MDLRNKSQKWKWNRITLSSLEFALTGIFTAFKERNMKHVASALLAVIAGLVFKVSV 62

Query: 61 IEWLFLLLSIFLVITFEIVNSAIEINVVDLASDYHFSMLAKNAKMAAGAVLMISGAVLT 120
      IEWLFLLLSIFLVIT FEIVNSAIEINVVDLASDYHFSMLAKNAKMAAGAVL+ISG+A LT
55  Sbjct: 63 IEWLFLLLSIFLVITFEIVNSAIEINVVDLASDYHFSMLAKNAKMAAGAVLVISGFAALT 122

Query: 121 GLIIFPKIWNIFVH 135
      GLIIF+PKIW + H
      Sbjct: 123 GLIIFVFKIWFLPH 137
  
```

60 An alignment of the GAS and GBS proteins is shown below.

-2465-

Identities = 98/129 (75%), Positives = 115/129 (88%), Gaps = 2/129 (1%)

Query: 1 MDLNDNN--HKWQNRITLTSSMEFVGTGIFTAFKEERNRKHLSAILVILAGLTQVSM 58
 M LNDNN +KWNRTI+TSS+EFA+TG+FTAFKEERN+R HL+SA L +AGL F +S
 5 Sbjct: 1 MALHDNNTTKRWKWRITLTSSLSFALTGVFTAFKEERNLRSHLLSACLACVAGLFFSISA 60

Query: 59 VEWLFLLLSIFLVNITPEIINSATENVVVLASNYPSMLAKNAKMDAGAVLVVSLFAVLV 118
 +EWLFLLL+FLVIT EI+NSATENVVLAS+YHPSMLAKNAKMDAGAVL++S +AVL
 10 Sbjct: 61 IEWLFLLLAIFLVITLIVNSATENVVLASDYHPSMLAKNAKMDAGAVLMSIGVAVLT 120

Query: 119 GLIIFIPKI 127
 GLIIFIPKI
 15 Sbjct: 121 GLIIFIPKI 129

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2187

A DNA sequence (GBSx2304) was identified in *S.galactiae* <SEQ ID 6755> which encodes the amino acid sequence <SEQ ID 6756>. This protein is predicted to be GTPase Era (era). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.1871 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10017> which encodes amino acid sequence <SEQ ID 10018> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA41632 GB:AF072911 GTPase Era [Streptococcus pneumoniae]
 Identities = 273/299 (91%), Positives = 290/299 (96%)

35 Query: 16 MTFKSGFVAILGRPNVGKSTFLNHVMQKIALMSDKAQITRNKIMGIVYTETQIVFIDT 75
 MTFKSGFVAILGRPNVGKSTFLNHVMQKIALMSDKAQITRNKIMGIVYT E QIVFIDT
 Sbjct: 1 MTFKSGFVAILGRPNVGKSTFLNHVMQKIALMSDKAQITRNKIMGIVYTETQIVFIDT 60

Query: 76 PGIHKPKTALGDPVVSAYSTLREVETVLFNVPADKRGKGGDDMIERLKAQKIPVILVI 135
 40 PGIHKPKTALGDPVVSAYSTLREV+TVLFNVPAD RKGDDMIERLKAQK+PVILV+
 Sbjct: 61 PGIHKPKTALGDPVVSAYSTLREVETVLFNVPADKRGKGGDDMIERLKAQKIPVILV 120

Query: 136 NKIDKVVHPDQLLSQIDDFRSDMFKEVVPISALQGNVPTILKLTNLSBGPQYFFSDQ 195
 NKIDKVVHPDQL QIDDFR+QMDFK+VPISALQGNV L+ +L++NL+BGQYFF DQ
 45 Sbjct: 121 NKIDKVVHPDQLSQQIDDFRQMDFKVPIPSALQGNVRLVDILSENLSBGPQYFFSDQ 180

Query: 196 ITDEPERFLVSEMVREKVLHLTQCEVPSVAVVVSMSKRDEETDKVHIRATINVERDSQK 255
 ITDEPERFLVSEMVREKVLHLT++E+PESVAVVV+SMKRDEETDKVHIRATINVERDSQK
 Sbjct: 181 ITDEPERFLVSEMVREKVLHLTREETPESVAVVVSMSKRDEETDKVHIRATINVERDSQK 240

50 Query: 256 GIIIGKQGAMLKKGKGMARRDIEMLGDKVYLETWVKVKKWRDKKLADPGYNEKSEY 314
 GIIIGK GAMLKKGK GMARRDIEMLGDKV+LETWVKVKKWRDKKLADPGYNE+EY
 Sbjct: 241 GIIIGKQGAMLKKGKGMARRDIEMLGDKVLETWVKVKKWRDKKLADPGYNEKSEY 299

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6757> which encodes the amino acid sequence <SEQ ID 6758>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

-2466-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1088 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 295/297 (99%), Positives = 296/297 (99%)

10 Query: 18 FKSGFVAILGRPNVGKSTFLNHVMQKIAIMSDKAQTTNRKIMGIYTTETEQIVFIDTPG 77
 FKSGFVAILGRPNVGKSTFLNHVMQKIAIMSDKAQTTNRKIMGIYTTETEQIVFIDTPG
 Sbjct: 2 FKSGFVAILGRPNVGKSTFLNHVMQKIAIMSDKAQTTNRKIMGIYTTETEQIVFIDTPG 61

15 Query: 78 IHKPTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDMMIIRLKAAKIPVILVINK 137
 IHKPTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDMMIIRLKAAKIPVILVINK
 Sbjct: 62 IHKPTALGDFMVESAYSTLREVETVLFMVPADEKRGKGDMMIIRLKAAKIPVILVINK 121

20 Query: 138 IDKVHDPQLLEQIDDFRSQMDFKFVVPISALQGRNVPTLIKLLTDNLEEGFYFFPDQIT 197
 IDKVHDPQLLEQIDDFRSQMDFKFVVPISALQGRNVPTLIKLLTDNLEEGFYFFPDQIT
 Sbjct: 122 IDKVHDPQLLEQIDDFRSQMDFKFVVPISALQGRNVPTLIKLLTDNLEEGFYFFPDQIT 181

25 Query: 198 DHPERFLVSENVREKVLHLTQQEVPHSVAVVSVESMKRDEETDKVHIRATIMVERDSQKGI 257
 DHPERFLVSENVREKVLHLTQQEVPHSVAVVSVESMKRDEETDKVHIRATIMVERDSQKGI
 Sbjct: 182 DHPERFLVSENVREKVLHLTQQEVPHSVAVVSVESMKRDEETDKVHIRATIMVERDSQKGI 241

30 Query: 258 IIGKQGAMLKKTGMARRDIELMLGDKVYLETWVVKVKKWRDKLADFGYNEKEY 314
 IIGKQGAMLKKTGMARRDIELMLGDKVYLETWVVKVKKWRDKLADFGYNEKEY
 Sbjct: 242 IIGKQGAMLKKTGMARRDIELMLGDKVYLETWVVKVKKWRDKLADFGYNEKEY 298

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2188

A DNA sequence (GBSx2305) was identified in *S.galactiae* <SEQ ID 6759> which encodes the amino acid sequence <SEQ ID 6760>. Analysis of this protein sequence reveals the following:

35 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2679 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2189

A DNA sequence (GBSx2306) was identified in *S.galactiae* <SEQ ID 6761> which encodes the amino acid sequence <SEQ ID 6762>. Analysis of this protein sequence reveals the following:

50 Possible site: 21
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

-2467-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAA16793 GB:D90900 hypothetical protein [Synechocystis sp.]
Identities = 36/119 (30%), Positives = 57/119 (47%), Gaps = 15/119 (12%)
Query: 390 TSDYSKARKVIHDLVNNYATRELATTRETASGISIHAEALYKDKRGVCOAFVAMFKD 449
++D+S+A++ + N Y +A TR I PS + +C ++ ++
10 Sbjct: 153 SNDWSEARLAYSWITQNTAYDVP-MAETRN-----IDDLRPETVLARGETICSGYSNLQA 207
Query: 450 MAATAGLSVWVYTGQAGG-----NHANIVTINGVKYVDITWDRNIKSNIKYF 498
+A GL V + G A G S NHANV V I+G Y +DTTW I S+ F
15 Sbjct: 208 LAKEGLDGVIIIBGFAKSGDIVGDDPVDNHNWNGVKIDQWYLLDTTWAGIVSDGRF 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6763> which encodes the amino acid sequence <SEQ ID 6764>. Analysis of this protein sequence reveals the following:

Possible site: 23
20 >>> May be a lipoprotein
----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 41/181 (22%), Positives = 79/181 (42%), Gaps = 17/181 (9%)
30 Query: 355 ITITYTLKGDVGLHKEYKQFVDSFVKENITNNITSDYERAKVIHDLVNNYATYATE-- 412
+ +T+ + D +++ Q + + + N +K+ YR+ K + + + + Y +
Sbjct: 124 VFVTFPLPEDAKNIYQDL-QAIGNDIVANTPSKD---RYRQVKYFYEVIRDTYDNKGF 179
35 Query: 413 ELATTRETASGISIHAEALYKDKRGVCOAFVAMFKDMAATAGLSVWVYTGQAGG-- 469
E + A S +++ D VC +A F+ + AG+ V Y+ G
Sbjct: 180 EAYQSGSQAQVASNQDIKSVFIDHLSVCNQAQAQFLOQAGIFVAYIRGTOTSQQPQQ 239
40 Query: 470 ---HANNIVTINGVKYVDITW-----DNNIKSNKYFLVGKTIAMDHLDSQYNALAKDI 522
HANN V IN Y VD TW DN++ K + + + L + + + KDI
Sbjct: 240 SFAHANNVQINNTYTGVDVTDGDFVFNHLSHQKQTINYSFLCLPDYLMALSHQPKDI 300

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

45 Example 2190

A DNA sequence (GBSx2307) was identified in *S.agalactiae* <SEQ ID 6765> which encodes the amino acid sequence <SEQ ID 6766>. This protein is predicted to be rgg protein. Analysis of this protein sequence reveals the following:

Possible site: 29
50 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.16 Transmembrane 187 - 203 (187 - 203)
----- Final Results -----
bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-2468-

A related GBS nucleic acid sequence <SEQ ID 10015> which encodes amino acid sequence <SEQ ID 10016> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5 >GP:AAA26968 GB:M89776 rgg [Streptococcus gordonii]
   Identities = 71/273 (26%), Positives = 140/273 (51%), Gaps = 16/273 (5%)

Query: 8  KELGKTLRFLRNGKKVSISSLADEHLGSKQISRPFEGESRITCSRLNIILDKLANITIDRF 67
      K GK L+ +R+ K+S+ +A +S +Q+SR+ERG S+T +L +++++ EP
10 Sbjct: 5  KSSGKILKIITRESKMSLKVAAGDISVAQLSRYERGSISSLVDSFYSCLRHMSVSLAEF 64

Query: 66  VSI-HSKAHTHFILLNRVKYCAEKHVTKLAVL-----EDHNHKDYKIMIK 115
      +H+ +L ++ + E N+ KL +L+ E N+K I+I+
10 Sbjct: 65  QYVYHNYREADDVVLSQLSEAGQENNVKLESILAGSEMAQEPPEKIONYK-LNITVIR 123

Query: 116  ALPSIDQSIIEPNQEELARLTDYLFPTVQWGYEIIILGNCSRLINVTPLFLTKEMVNS 175
      A+S + + + + + L TDYLF+VE+WG YE+ L N L+ TL EM+N
15 Sbjct: 124  ATLTSNPDYQVSKGDIETLTDYLPVSEWGRYELMLPTNSVNLTLTLETFASEMINR 183

Query: 176  FAYSEQNRKNIKILVTQLAINCLIIISIDHSYPESHYILDKVRSLLQDEVNPFYKTVFLVY 235
      + N+ ++ ++ +N+ I++++ + + + + + R+ Y++ + Y
20 Sbjct: 184  TQFYNNLPENRRIIIMLLARVVSACIENSHLQVAMKPLNYIDNTKIPETOLYDRVLKIKH 243

Query: 236  TGYHHLKGLDTSKGDMDRKLQIPKYLGEDSF 268
      Y K+G+ + + D+ + L F+YL DSF
25 Sbjct: 244  KALYSYKVGMPHA-RHDIEQCLSTFPEYL--DSF 273

```

There is also homology to SEQ ID 628.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2191

A DNA sequence (GBSx2308) was identified in *S. agalactiae* <SEQ ID 6767> which encodes the amino acid sequence <SEQ ID 6768>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3234 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA05066 GB:D26071 formamidopyrimidine-DNA glycosylase
   [Streptococcus mutans]
   Identities = 182/271 (67%), Positives = 217/271 (79%)

45 Query: 1  MPFLPEVETVRKGLERLVANQRIASITIKVPKRVKTDLMDFMISLPGKTIQQVLRKGYL 60
      MPFLPEVETVR+GLE L+V ++I S+ ++VPRKVKT + DF + + G+T + + RKGKYL
50 Sbjct: 1  MPFLPEVETVRKGLHLTVGKKIVSVVRVPRKRVKTVGDFQILDILQQTTFESIGRGGKYL 60

Query: 61  LFDPGSMVMVSHLRMEGKYLLFPNPKV/PNKHFLHYFKL/NGSTLVYQDVRKFGFTFLVRK 120
      L + ++SHLRMEGKYLLF ++VPRKHFHL+ F L GSTLVYQDVRKFGFTFL+ K
50 Sbjct: 61  LMLNLRQTIIISLRMEGKYLLFEDVRPDKRHFHLPFGLDGGSTLVYQDVRKFGFTFL+PK 120

Query: 121  SSLKYDTFTQKKLGPETADTFQFEPFSKGLANSKRPKPLLLDQRLVAGLGNITVDEVLM 180
      S + + YF QKK+GPEP A F+ +PF +GLA S K LK LLLDQ LVAGLGNITVDEVLM
55 Sbjct: 121  SQVYATFVQKKIGPEPNKADFKLDPFEBGLAKSHVITKLLLDQHLVAGLGNITVDEVLM 180

Query: 181  AAXIHQRLANQLTESETSLLEKHEIRILTLGIBKOGSTIRTYKNALGEDGTMKYLVY 240

```

-2469-

AAK+ P+RLA+QL SE +H E IRIL L IEKGGSTIR+YKN+LGEDG+MQ LQVY
 Sbjct: 181 AAKVDPERLASQLKTSSEIKRIHDETRIRILQLAIEKGGSTIRSYKNSLGEDGMQCLQVY 240

Query: 241 GKTGQPCPCGCLIKKIKVGGRGTHYCPRCQ 271
 GKT GPC RC I+KIKVGGRGTH+CP CQ
 Sbjct: 241 GKTGQPCARCATPPIKIKVGGRGTHFCPCSQ 271

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6769> which encodes the amino acid sequence <SEQ ID 6770>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2068 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 190/271 (70%), Positives = 229/271 (84%)

Query: 1 MFELPEVETVRKGLERLVNQLIASITIKVPMVKITLNDPMISLPGKTIQQVLRGKYL 60
 MFELPEVETVR+GLE LV+ QEI ++T+KVPIMVKITL F ++LPG+ IQ V RRGKYL
 Sbjct: 1 MFELPEVETVRGLLETVLVLSQEIIVAVTLKVPIMVKITOLETALITLPGQIQSVGRGKYL 60

Query: 61 LFDGEMVWVSHLRMEGKYLFPFNKVPDNKHFHLYFLKNGSTLVYQDVRKFGTFELVRK 120
 L D G++V+VSHLRMEGKYLFP++VPDNKHFH++F+L NGSTLVYQDVRKFGTF+L+ K
 Sbjct: 61 LIDLQQLVLVSHLRMEGKYLFPDDEVFDNKHFRHVFPELNGSTLVYQDVRKFGTFLLDAK 120

Query: 121 SSLKDYFTQKKLGPEPTADTFQEPFSSKGLANSKKPIKLLLOQLVAGLGNIYDEVLM 180
 S L +F ++KLGPFP +TF+ + F L +S+KPIK LLDQ LVAGLGNIYDEVLM
 Sbjct: 121 SQLSAFFAKRKLGPFPKATTFKLKTFEALLSSQKPIKHLDDQLVAGLGNIYDEVLM 180

Query: 181 AAKIHPRLANQLTSESTSLHKBIIRILTIGIEKGGSTIRTYKNALGEDGTMQKYLQVY 240
 AAK+HP+ ++L ++E LH E IRIL LGIEKGGST+RTY+NALG DGTGM YLQVY
 Sbjct: 181 AAKVHPEFASRLAKAEIKRLHDETRIRILALGIEKGGSTVRTYRNALGADGTWJDYLQVY 240

Query: 241 GKTGQPCPCGCLIKKIKVGGRGTHYCPRCQ 271
 G+TG+PCPCG I K+KVGGRGTH CP+CQ
 Sbjct: 241 GGTGKPCPCGQAIVKLKVGGRGTHICPCRCQ 271

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2192

A DNA sequence (GBSx2309) was identified in *S.agalactiae* <SEQ ID 6771> which encodes the amino acid sequence <SEQ ID 6772>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0797 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10013> which encodes amino acid sequence <SEQ ID 10014> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA00353 GB:AF008220 YtaG [Bacillus subtilis]

-2470-

Identities = 80/189 (42%), Positives = 113/189 (59%), Gaps = 1/189 (0%)

Query: 8 MTKIIGLTGGIAGSGKSTVTKIIRSGPKVIDADQVVHKLQAGGKLYQALLEWLGFEILD 67
 5 Sbjct: 1 MTIIGLTGGIAGSGKSTVAMLMIEGKITVIDADITAKQAVEKGMFAYRQIIDEFGEDLL 60

Query: 68 ADGELDRPKLSQMIFANFDMKTSARLQNSIIRQELACQRDQLKQTEEIF-FMDIPELLIE 126
 ++G+DR KL ++F N ++ +RQE+ +RD+ E F +DIELL E
 10 Sbjct: 61 SNGDIDRKKLGALVPTNEQKRLALNAIVHPAVRQEMLNKRDEAVANREAPVVLIDIELLFE 120

Query: 127 EKYIKWFDRIWLVPVDKEKQLQRLMARNYSRREAEIRLSHQMPLTDKKSFASLIIDNNG 186
 K D+I +V V KE QL+RLM RN +EEA R+ QMEL +K +A +IDN+G
 15 Sbjct: 121 SKLSSLVDKIIIVSVTKELQLERLMKRNQLTBEAEVRSIRSQMPELEKRTARADQVIDNSG 180

Query: 187 DLITLKEQI 195
 L K Q+
 15 Sbjct: 181 TLEETKRQL 189

A related sequence was also identified in GAS <SEQ ID 9111> which encodes the amino acid sequence
 20 <SEQ ID 9112>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty= 0.101(Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 118/191 (61%), Positives = 153/191 (79%)

Query: 9 TKIIGLTGGIAGSGKSTVTKIIRSGPKVIDADQVVHKLQAGGKLYQALLEWLGFEILDA 68
 T IIG+TGGIAGSGKSTV K+IR++G++VIDADQVVH LQ KGG+LY+AL E G +IL A
 35 Sbjct: 9 TMIIGLTGGIAGSGKSTVVKVIRKAGYQVIDADQVVHDLQKGGRLYKALREAFGNQILKA 68

Query: 69 DGELODRPKLSQMIFANFDMKTSARLQNSIIRQELACQRDQLKQTEEIFMDIPELLIEBK 128
 DGBLDR KLS-N+F+NFERN TS+ +QN II++ELA +RD L Q++ IFMDIPLG+E
 35 Sbjct: 69 DGBLDRTKLSEHLFSNPNMATSSAIQNQIIEELAAKEDHLAQSAIFMDIPLMLG 128

Query: 129 YINWFDEIWLVPVDKEKQLQRLMARNYSRREAEIRLSHQMPLTDKKSFASLIIDNNGDL 188
 Y NFD IWL+VD +QLQRLMARN + +A R++ Q+P+ +KK +ASL+IDN+GD+
 40 Sbjct: 129 YQWFDADWLVPVVDATQLQRLMARNRLKGGKARQIASQLPIEEKFPYASLVIDNSGDI 188

Query: 189 ITLKEQILDAL 199
 L +Q+ AL
 45 Sbjct: 189 AALIKVQSAL 199

A related GBS gene <SEQ ID 8993> and protein <SEQ ID 8994> were also identified. Analysis of this
 protein sequence reveals a signal peptide at residues 1-16.

50 The protein has homology with the following sequences in the databases:

42.2/60.6% over 189aa

OMNI|NT01BS3382| Insert characterized

55 ORF02237(319 - 885 of 1206)
 OMNI|NT01BS3382(3 - 192 of 200) ()
 %Match = 17.0
 %Identity = 42.1 %Similarity = 60.5
 Matches = 80 Mismatches = 74 Conservative Sub.s = 35

60 78 108 138 168 198 228 258 288
 KNSPTAFG*SIDRI*NKLTITQNYSHFNFRHRRKWLJD*NI*ECSWRGRYDAKVFTGLN*NGWATVSKVLEFN*EDKSRRE

-2472-

transport protein [Pyrococcus horikoshii]
 Identities = 78/240 (32%), Positives = 130/240 (53%), Gaps = 13/240 (5%)

Query: 3 LVIRDIRKRQETEVLRGASVRYFSGKITVGLRNGAGKITLFNIIYGLDIAADNGKITLL 62
 +++ ++RK+F EVL+G ++ G+I G+LG NG+GK+T IL G + G + +
 Sb|ct: 2 IIVENLRKKFGSKVGLGINTVNDGETYGLLGPNKSGKSTTWRLSGIITDFGKVMVA 61

Query: 63 -KINHEYPIATKDI-GIVISNYLPEFLTGYPFKFYMDLI--PSDDL-MTIDDYLDME 117
 D P+ K+I G V L E L T E F F + P D L + +D
 Sb|ct: 62 GVDVSRDPMKVKEIVGVVPEPALVESLTPAEPPSFIGGVRRIPQDILERNKRLVDAPG 121

Query: 118 IQQTERHRLIKGYSDMKSKLSLICLMISKPKVILLDRFLTAVDVVSSIAIKRLRLISE 177
 IG+ +++I S G K K+SLI ++ P+V++LDS + +D S+ + L E E F
 Sb|ct: 122 IGR-YMQLIGLTLSPFTKQKISLISALLHDPQLILDERMNGLDPKSARIPRELLFEPKE 190

Query: 178 D-HIILSTHIMALAEDLCDIVAVLDKGL---QTLDDIR--KHEQFEERLLVLKGD 230
 + I+ STHI+ALAE +CD + ++ G++ T+D R + E+ E+ L++ + E
 Sb|ct: 181 EGKSI VFS THILALAEVACRIGIITYGRI VASGTIDELREIAREEKLIDFLKLTAKE 240

20 There is also homology to SEQ ID 2876.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2194

A DNA sequence (GBSx2311) was identified in *S. agalactiae* <SEQ ID 6777> which encodes the amino acid sequence <SEQ ID 6778>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.6138 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2195

A DNA sequence (GBSx2312) was identified in *S. agalactiae* <SEQ ID 6779> which encodes the amino acid sequence <SEQ ID 6780>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -15.34	Transmembrane	526 - 542 (511 - 546)
INTEGRAL	Likelihood = -9.61	Transmembrane	340 - 356 (335 - 359)
INTEGRAL	Likelihood = -8.17	Transmembrane	455 - 471 (451 - 476)
INTEGRAL	Likelihood = -8.01	Transmembrane	97 - 113 (95 - 121)
INTEGRAL	Likelihood = -8.01	Transmembrane	216 - 232 (207 - 236)
INTEGRAL	Likelihood = -3.40	Transmembrane	50 - 66 (46 - 67)
INTEGRAL	Likelihood = -1.33	Transmembrane	178 - 194 (178 - 194)

----- Final Results -----

bacterial membrane	---	Certainty=0.7135 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

-2473-

A related GBS nucleic acid sequence <SEQ ID 10011> which encodes amino acid sequence <SEQ ID 10012> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 376.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2196

10 A DNA sequence (GBSx2314) was identified in *S.galactiae* <SEQ ID 6781> which encodes the amino acid sequence <SEQ ID 6782>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence
15  INTEGRAL  Likelihood = -8.17  Transmembrane  140 - 156 ( 134 - 160)
    INTEGRAL  Likelihood = -6.64  Transmembrane  255 - 271 ( 253 - 274)
    INTEGRAL  Likelihood = -5.79  Transmembrane  345 - 361 ( 343 - 363)
    INTEGRAL  Likelihood = -3.29  Transmembrane  184 - 200 ( 183 - 202)
    INTEGRAL  Likelihood = -2.34  Transmembrane  66 - 82 ( 65 - 83)
    INTEGRAL  Likelihood = -1.65  Transmembrane  221 - 237 ( 221 - 239)
20  INTEGRAL  Likelihood = -0.00  Transmembrane  121 - 137 ( 121 - 137)

----- Final Results -----
      bacterial membrane --- Certainty=0.4270 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9401> which encodes amino acid sequence <SEQ ID 9402> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:CAA07482 GB:AJ007367 multi-drug resistance efflux pump
    [Streptococcus pneumoniae]
    Identities = 213/372 (57%), Positives = 295/372 (79%)

35  Query: 1  MPPMVLXVEQLGAPSNKVENWYAGLSVSLGALGSSALVAPLAWRIADRYGRKPMWVRAGLAM 60
      +PFM +VVE LG S +V +YAGL++S+SA+S+AL +P+WG LADRYGRKPMW+RAGL M
      Sbjct: 28  VPFMPFIVENLVGSGQVAFYAGLAISSAISAALPSPINWILLADRYGRKPMWIRAGLAM 87

      Query: 61  TTFMGGLAFHSHVTTGILLIRILINGIPAGYVPNSTALIASQAPQESGYALGTIATGVVGG 120
      T TMGGLAF+ ++ L+ LR+LNG+PAG+VPM+TALIASQ P+ESG ALGT+TGV G
40  Sbjct: 86  TITMGGLAFVPMIYWLIFLLRLNGVPAGVPNNATLIASQVQPKKSSGALGTLTGCVVGG 147

      Query: 121  MLIGFLHGGILAFWFGIREVFLNKGITLITLMTIIPWKKDPKVISNKKIMPTTEVVKSS 180
      L GP +GG +AE FGIR VFLVGG L ++ ++TI +KED+P++ E+ +PT E+E S
45  Sbjct: 148  TLTGPFITGGFIARLFGIRTVFLVGGFLFLAAILTICFIKEDPQFVAKEAIPITKELPTS 207

      Query: 181  VKSLQILIGLFTVSMIIQISAQSTAPITLTYIRHLQQTENLMPVSGLIVSGMGFSSILES 240
      VK +L+ LP+TS +IQ SAQSI PTL LY+R LQQTENL+PVSGLIVS MGFS+AS+
50  Sbjct: 208  VKYFYLILNLFITSPVICFSAQSIGFITALYVRLDQQTENLFPVSGLIVSSMGFSSMSA 267

      Query: 241  PKLGRIGDRIGNHRLLALLLAYSFLMYVCLSAQTSLQLGVIRFLYGPOTGALPMSINSI 300
      +G+GD+GNHRLL+A YS ++Y+LC+ A +LQGA+ RFL+G GTGAL+P +N+
      Sbjct: 268  GVMGKLGDKVGNHRLLVVAQFYSVIYILCANASSPIQLGLYRFLGLATGALIPVNAI 327

      Query: 301  LTKIAPRQGLSRIPSYNQPMNLGQVLGPFVGBAVSIHLGFRWVFVFTSFIVLANFVWCF 360
      L+K+ P+ G+SR+P++NQ+P LG V+GP GBVAV+ G+ VF+ TS V + ++

```

-2474-

Sbjct: 328 LSKMTFKAGISRVFAFMQVFFYLGGVVGPMAGSAVAGQFGYHAVFYATSLCVAFSCLENL 387

Query: 361 INFRYIRVKEI 372

I FR ++VKEI

5 Sbjct: 388 IQFTLLKVKRI 399

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6783> which encodes the amino acid sequence <SEQ ID 6784>. Analysis of this protein sequence reveals the following:

Possible site: 58
 10 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -10.14 Transmembrane 165 - 181 (150 - 185)
 INTEGRAL Likelihood = -7.43 Transmembrane 371 - 387 (367 - 391)
 INTEGRAL Likelihood = -3.88 Transmembrane 90 - 106 (86 - 109)
 INTEGRAL Likelihood = -3.35 Transmembrane 145 - 161 (143 - 162)
 15 INTEGRAL Likelihood = -1.70 Transmembrane 279 - 295 (279 - 297)
 INTEGRAL Likelihood = -0.85 Transmembrane 209 - 225 (209 - 226)
 INTEGRAL Likelihood = -0.27 Transmembrane 347 - 363 (347 - 363)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.5055 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:CAA07482 GB:AJ007367 multi-drug resistance efflux pump
 [Streptococcus pneumoniae]
 Identities = 236/396 (59%), Positives = 309/396 (77%)
 30 Query: 1 VNNRQNLKVAWLGSPFTGASFLVMPFMALYVENLGTPTLVEYVYAGLAVAVTALASALF 60
 ++N+ NL++AW GNF TGAS SLV+PFM ++VENLG ++ V +YAGLA++V+A+++ALF
 Sbjct: 4 INWKNLRLIAWFGNFTGASISLVVFPMTFVENLGVGSGQVAFYAGLALSVAISALASALF 63
 Query: 61 APWKGKLADRYGRKPMMLRASVPMFTFMGGLAIPNPFVLLILRLLTGVSAQYVVPATL 120
 +P+WG LAD+YGRKPMH+RA MT TMGGLA +PM++NL+ LRL GV AG+VVPATL
 35 Sbjct: 64 SPIWGILLADRYGRKPMMLRAGLMTITMGGLAFVPIYNI LRLINGVFAQVVPATL 123
 Query: 121 IASQAPKERSGYALGTLTGTAGALIGPLLGGILAEILGIRQVLLVGVILFGLSMTA 180
 IASQ FKE+SG ALGTL+TGV AG L GP +GG +AEL GIR VELLVG LFL +++T
 40 Sbjct: 124 IASQVPEKSGSALGTLTGTGVAGTLTGPFPGFIAELFGIRTVFLLVGSFLFLAAILITI 183
 Query: 181 VYVKEKPKVTRFEMHPTKVLKVGKSPQIMLGLFPTSMIIQISASQVAPILSLYIRHL 240
 ++KE+P+PV ++ IPTK + VK P ++L LE+TS +IQ SQAQ+ PIL+LYAR LG
 Sbjct: 184 CFIKEDFPQVAKKAIPTELFTSVKTYPLLNLPLTSPVIGQSAQSIGPTILAYVRDLG 243
 45 Query: 241 QTNLLMPTSLGVASVAMGFSSFLPSSYLCKIGRPFQWRILLALCYSTFMVSSALACTS 300
 QT NL+P SGL+V8+MGFSS+ S+ +OKLQD+ QNRLL+ A YS I+Y A A +
 Sbjct: 244 QTNLLPVSLIVSSMGFSSMWSAGVWGLQDKVGNRLLLVVAQFYSVIIYLLQASASP 303
 50 Query: 301 FQLGVLRFAYQFGVGLMPSINSLLTKLTPKEGISRVPAYNQWPSNLQCVIOPFTGSNVA 360
 QLQ+ RF +G G QAL+P +N+LL+K+TEK GISRVFA+NQ+P LG V+GP GS VA
 Sbjct: 304 LQLGLYRFPLGLGTGTLIPGVNALLSKMTFKAGISRVFAFMQVFFYLGGVVGPMAGSAVA 363
 Query: 361 VVLGYSRVFVTVSLIVFNILNSLIIFPKYIKVDI 396
 GY +VEY TSL V + +++LI FR +VK+I
 55 Sbjct: 364 QQFGYHAVFYATSLCVAFSCLENLQFTLLKVKRI 399

An alignment of the GAS and GBS proteins is shown below.

Identities = 262/373 (70%), Positives = 314/373 (83%)

60 Query: 1 MPFMVLYVEQLGAPENKVEWYAGLSVLSLSSALVAPLWGRGLADRYGRKPMVVRGLAM 60
 MPFM LYVE LG P+ VE+YAGL+V+++AL+SAL AP+WG+LAD+YGRKPM+RA +M
 Sbjct: 25 MPFMALYVENLGTPTLVEYVYAGLAVAVTALASALFAPWVGKLADRYGRKPMMLRASVPM 84
 Query: 61 TFTMGSLAFIHVSITGLLILRLINGIFAGYVNPSTALIASQAPQERSGYALGTLTGTG 120

-2475-

TPTMGGLA I +V LLILR+L G+ AGYVFN+TALIASQAP+RESGYALGTIATGVT G
 Subjct: 85 TPTMGGLAIIPIHVFLLILRLITGVSAGYVFNATALIASQAPKEBSGYALGTIATGVTAG 144
 Query: 121 MLIGPLGLGLLAEKFGIREVFLLAGTILLISTIMTIPMKEDFKPISNEEIMPTTEVFKS 180
 LIGPLGLGG+LAE GIR+VPLNG IL + +LMT VKE+PK+ E +PT + K
 Subjct: 145 ALIGPLGLGLLAKLIGIQVFLNGVILFLCSLMTAVYKKEFKPVRRFEMIPKTVILQG 204
 Query: 181 VKSLQILIGLGVNMIQISNQSIAPIILTYIRHLAQTNLMFVSGLVSGMGFSSILSS 240
 VKS QI++GLFVTSMIQISNQS+APIL+LYIRHLAQTNLMF SGL+VS MGFSS+ SS
 Subjct: 205 VKSPQIMLGLFVTSMIQISNQS+VAPILSLYIRHLAQTNLMFTSGLVVSAMGFSSLFSS 264
 Query: 241 PKLGRIGTRIGNHRLLLALLYSFIMYVCLSAQTSQGLGVIRPLYGPGTALMPSINSI 300
 LG++GDR GNRHLL AL YSF+MY +LAQTS QIGV+RF YGFG GALMPSINS+
 Subjct: 265 SYLGLGDRPGNRHLLAALCYFSIMYFSALAAQTSFQLGVLRPAYGFGVGLMPSINSI 324
 Query: 301 LTKIAPFQGLGRIFSYNOMFSLGQVLPFFVGSVAISHLGFERNVFTTSPIVLANFVQCF 360
 LTK+ P+G+SR+P+YKMGFSLGQV+GPF+GS V+ LG+R VF+VTS IV N +W
 Subjct: 325 LTKLTFKEGISKVFAYNOMFSLGQVIGPFVGSNAVVLGYSRVFVTSILVFNLIWSL 384
 Query: 361 INFRKYIRVKEIV 373
 I FRKYI+VK+IV
 Subjct: 385 IIFRKYIKVKDIY 397

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2197

A DNA sequence (GBSx2315) was identified in *S. agalactiae* <SEQ ID 6785> which encodes the amino acid sequence <SEQ ID 6786>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2343 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB69986 GB:U94356 glycerol kinase [Enterococcus faecalis]
 Identities = 156/186 (83%), Positives = 167/186 (88%), Gaps = 1/186 (0%)
 Query: 3 SEEKIYMAIDQGTSSRAIIPNKKGEKIASQKEFPQIFPQAGVVEHNAQIWNVSQSVI 62
 +SEKIYMAIDQGTSSRAIIP+KKG KI SQKEF Q FP AGVVEHNA+IWNVSQSVI
 Subjct: 2 AEEKIYMAIDQGTSSRAIIPKKNIGKSQKEFTQYFPNAGVVEHNAEIMNSVQSVI 61
 Query: 63 AGAFTBSIKPQDIEAIGITMRETTVVDKKTGLPIYNAIVWQSRCTAPIADQLKQKH 122
 AG+ IRS +KP I IGTIMRETTVVDK TGLPIYNAIVWQSRCT PIADQLK+G+
 Subjct: 62 AGLSLBSGVKPTDIAGIGITMRETTVVDKATGLPIYNAIVWQSRCTPIADQLKEDGY 121
 Query: 123 TMIHEKTLGLVIDAYPSATKVRWILDHVPGAQERAKGELLPGTIDTWLWKLTDGIVHV 182
 + MIHEKTLG+IDAYPSATKVRWILDHV GACERAE GRL+PGTIDTWLWKLTD G HV
 Subjct: 122 SEMIHEKTLGLIDAYPSATKVRWILDHVPGACERANKEMLPGTIDTWLWKLTDGLTHV 180
 Query: 183 TDYSNA 188
 TDYSNA
 Subjct: 181 TDYSNA 186

There is also high homology to SEQ ID 2844:

Identities = 174/186 (93%), Positives = 182/186 (97%)
 Query: 3 SEEKIYMAIDQGTSSRAIIPNKKGEKIASQKEFPQIFPQAGVVEHNAQIWNVSQSVI 62
 S+EKIYMAIDQGTSSRAIIPN+KKEK+SQKEFPQIFP AGVVEHNAQIWNVSQSVI

-2476-

- Sbjct: 2 SQEKYIMADQGITSSRAIIPNQKGEKVSSSQKEFPQIFPHAGNVSHNANQIWNVSQSVI 61
- Query: 63 AGAFIASSIKPGQITRAIGITNQRRTTVWDDKKTGLPIYNAIVMQSRQTAPIADQLQKQGH 122
AGAFIASSIKP QIRAITGINTQRRTTVWDDKKTG+PIYNAIVMQSRQTAPIA+QLKQ+GH
- 5 Sbjct: 62 AGAFIASSIKPGQITRAIGITNQRRTTVWDDKKTGVPYNAIVMQSRQTAPIAEQLQKQGH 121
- Query: 123 TNNIHEKTGLVIDAYFSATKRWILHVPQAQERARKGELLPGTIDTWLVWKLTDGLVHV 182
T MIHEKTGLVIDAYFSATKRWILHVPQAQERARKGELLPGTIDTWLVWKLTDG VHV
- 10 Sbjct: 122 TNNIHEKTGLVIDAYFSATKRWILHVPQAQERARKGELLPGTIDTWLVWKLTDGAVHV 181
- Query: 183 TDYSNA 188
TDYSNA
- Sbjct: 182 TDYSNA 187
- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2198

A DNA sequence (GBSx2317) was identified in *S.agalactiae* <SEQ ID 6787> which encodes the amino acid sequence <SEQ ID 6788>. This protein is predicted to be glycyl-tRNA synthetase beta chain (glyS).

- 20 Analysis of this protein sequence reveals the following:

Possible site: 18
>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2933 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 30 >GP:CAB14468 GB:299117 glycyl-tRNA synthetase (beta subunit)
[Bacillus subtilis]
Identities = 315/687 (45%), Positives = 447/687 (64%), Gaps = 21/687 (3%)
- Query: 3 KDILLRLGLEELPAIVVTPSEKQLQOKNVKFLSHRLSPETVOIPSTPRRLAVRVKGLAD 62
+DILLRLGLEELPA + S QLG K + I++ ++ V++F+PRRLAV VK +A+
- 35 Sbjct: 4 QDILLRLGLEELPAIVVTPSEKQLQOKNVKFLSHRLSPETVOIPSTPRRLAVRVKGLAD 63
- Query: 63 QOTDLTDFKGPSKKIALDABGNFSKAQGFVZKGLSVDDIEPREVKGKEYVYVTKHET 122
+Q D+ E+ KGP+KKIALDRA+GN+KAA GF +G+G +V+D+ +EVKG EYV+V K +
- 40 Sbjct: 64 QQDDIKEAKGPAKKIALDADQNWTKAIGSPSKGQANVEDLYIKVKGIEYVVPQKFGA 123
- Query: 123 GKSAIDVLASVTEVLTELTPFPVNMHWBNSPEYIRPHTLVVLDDQALELDFLDHSGR 182
G+ +L ++ ++T L FP NM WN YIRP+ +V L + + SR
- 45 Sbjct: 124 GQETKSLLELGLITSLHFPKNNRWGHEIDLYIRPIKNIVALPGQDVIPFSTINWESGR 183
- Query: 183 ISRGHRFLGSDTRISSASSVEDDILQQFVIADMKPQIMVNOIHAIIEKKNIISVIDED 242
+GHRFLG + I S S+YE+ L+ Q V IAD R+QMI +Q+ + + N S+ +DED
- 50 Sbjct: 184 TTQGHRLFLGHEVSIESPATYERQLKQHVIAIDPSVRKQMQISLETMAAKNHWISVIDED 243
- Query: 243 LLNEVLNWEYVETAFPLGSFDEKYLDPREEVLVTSKNNHQRYFVVRDRDGKLLPNPISVRN 302
LL+EV +LVEYPTA GSF+ +L +PEEVLT+MK HORYF V+D+G LLP+FI+VRN
- 55 Sbjct: 244 LLDEVNHLVRYPTALYGSFSEFSLIPREVLVITMKHQRYPFVKDKNGDLLPHFITVRN 303
- Query: 303 GNAEHNENVIKNEKVLVARLEDEGFPOEDQKMIADIVLEKLVKOVTFHKKIGSLGLEYHD 362
GN+ IENV +GNEKVL ARL D FP+ +EDQKMI V+KI+ + FHE+AGSL +
- Sbjct: 304 GNSHAIVENARGNEKVLVARLEDSAPFYKEDQKMIADANVKKLENTVHEELGSLADKVR 363
- Query: 363 RVKVIQYLAIEKADLSDEKIALVRAASTYKPOLITGMVDEFDELQGMKEYYALLAGRD 422
RV I++ LA + + + V RAA I KPOL+T M+ EF LQGMKEYA+ GE
- 60 Sbjct: 364 RVTSIAEKLAIVLQADEDTLKHVKRAASTISKPOLVTHMIYEPPELQGMKEYYARMLGED 423

-2477-

Query: 423 PAVAAAIKSHYMPTSDGELPSTRVGAIIAIAADKFDTLISFFSVGLIPSGSNDPYALRR 482
 A/AAA+ EHYMP SA GE P T GA++A+ADK DT+ SFPS+G+IP+GS DEF L R
 Sbjct: 424 EAVAAVNEHYMPTSDGELPSTRVGAIIAIAADKFDTLISFFSVGLIPSGSNDPYALRR 483

5 Query: 483 TQGVIRILEAFGWDIPDELVTNLYGLSFASLDYANQKVFAPISARIKEMIGS -KVPKD 541
 GIV IL W I +EL+T F D N E++ F +R++ ++ + ++ D
 Sbjct: 484 ASGVAILLRDMGISEFELLT-----FVQIDKEN -ELLDFFTORLKYVIMAEQIRHD 535

10 Query: 542 IREAVLESDDTYIVSLILEASQALVQKSKDAQYKVSVESLSKAFNLAEKVTHSVLVDSSLF 601
 + +AVLES L +Q L CK +K +E+L R +++++K + LF
 Sbjct: 536 VIDAVLESSELEPYALSHAKQAQLKQKLGAPGFKETAEALGRVISISKNGVRGD -IQPDLF 594

Query: 602 ENNQKALYQAILSLLELTEDMDNLDK-----LFAISPIINDFFDNTMVMVTDDEKM 652
 EN E L+ A + + E++ +N K L AL I+ +FD+TMV+ D+E +
 15 Sbjct: 595 ENEYAKLFDAYCTAK -ENLQENPSKKDYEAALASLAALKSEFDAYDHTMVIADNLSL 652

Query: 653 KQNRILAIIINSLVAKARTVAAPNLINTK 679
 K NRILA + SL + ++ A N L K
 20 Sbjct: 653 KNRILAQMSLADEIKSPANMALIVK 679

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2835> which encodes the amino acid sequence <SEQ ID 2836>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -0.96 Transmembrane 450 - 466 (450 - 466)

----- Final Results -----
 bacterial membrane --- Certainty=0.1383 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 505/679 (74%), Positives = 578/679 (84%)

35 Query: 1 MYDQLLELGLLELPAYVVTPEKQLGQRMVLFEDHRLSPSTVQIFSTPRRLAVRVGL 60
 M+K+LL+ELGLLELPAYVVTPEKQLG++ FL ++RLSFE +Q PSTPRRLAVRV GL
 Sbjct: 1 MSKNLLLELGLLELPAYVVTPEKQLGERLATFLTEHRLSPSTVQIFSTPRRLAVRVSGL 60

40 Query: 61 ADQQDILTDFKGPCKKIALDAGNFSKAAQGFVRGKGLSVDDIEFRVKGSEYVVTYKH 120
 ADQQDILTDFKGP+KICIALDA+GNFSKAAQGFVRGKGL+ D IFRFRVKGSEYVVTYKH
 Sbjct: 61 ADQQDILTDFKGPCKKIALDAGNFSKAAQGFVRGKGLTDAIEFRFRVKGSEYVVTYKH 120

Query: 121 ETGKSAIDVLASVTELTETFPVNMHMANNSFEYIRPVHTLVLLDDQALELDFLDHS 180
 E GK A +VL VTEVL+ +TFPV+HMANNSFEYIRPVHTL VLL+D+ALELDFLDHS
 45 Sbjct: 121 EAGKFAKEVLLGVTELSAMTFPVMHMANNSFEYIRPVHTLVLLDQALELDFLDHS 180

Query: 181 GRISGRHFLGSDTEISSASSYEDDLRQOFVIADAKERQMIWQIHAISEKKMINSVID 240
 GR+SGRHRFLG+T I+SA SYE DLR QFVIADAKERQ+MIV QI ++ ++ + +V+ID
 50 Sbjct: 181 GRVSGRHRFLGTETTTTSDSYEADLRSQFVIADAKERQMIWECIKTLEVSQGVQVQID 240

Query: 241 EDLLANEVNLVPEYPTAFGLSGFDEKYLDFVESVLVTSKMKHGRYFVVRDQGLKLPNFSV 300
 EDLLANEVNLVE+PTAF+GSF+ KYLDFVEEVLVTSKMKHGRYFVVRD+ G L+ENF+SV
 Sbjct: 241 EDLLANEVNLVPEYPTAFGLSFEAKYLDFVESVLVTSKMKHGRYFVVRDQGLHLMFVFSV 300

55 Query: 301 RKNQAEHIENVIKNGKVLVARLEDGEFFWQEDQKLMIALDVEKLQVTFHEKIGSLVH 360
 RKNQ + IENVIKNGKVLVARLEDGEFFW+EDQKL IADVL KL VTFHEKIGSL EH
 Sbjct: 301 RKNQDAIENVIKNGKVLVARLEDGEFFWQEDQKL IADVLAKLTNTVTFHEKIGSLAHK 360

60 Query: 361 MDRVVISQYLAEKADLGESEKLAIVLRAASIVKOLLTQWDFDLQGIIMGEKYALLAG 420
 MDR +VI+ LA++A+LS EE AV RNA IYKFLTLTMY EDELQGIIMGEKYALLAG
 Sbjct: 361 MDRITVLAASLAKAENLSAREVTVADRAQIYKFLTLTQWDFDLQGIIMGEKYALLAG 420

Query: 421 EQPAVAAAIRESHYMPTSDGELPSTRVGAIIAIAADKFDTLISFFSVGLIPSGSNDPYALR 480
 E AVA AIREHY+P +A G LPRT+GGA+LALA K DTLISFFSVGLIPSGSNDPYALR
 65 Sbjct: 421 EDPAVATAIREHYLPDANGGALPSTRVGAIIAIAADKFDTLISFFSVGLIPSGSNDPYALR 480

-2478-

Query: 481 RATQGIIVRILEARFGWDIPLDELAVTNLYGLSFASLDYANQKEVMFISARIEKMTGSKVPK 540
 RATQGIIVRIL- PGM IP+D+LV +LY L6F SL YAN+ +VM FI AR++KM+G PK
 Sbjct: 481 RATQGIIVRILDHFGWRIPMDKIVDSLYDLSFDSLTYANKADVMNPIRARVDKM+GKAAPK 540

Query: 541 DIREAVLESDDTYIVSLILRASQALPQESKDAQYKVSVESLSRAFNLAARKVTHSVLVDSLL 600
 DIREA+L S T++V +L A++ALV+ S YK +VESLSRAFNLAER SV VD SL
 Sbjct: 541 DIREAILASTSTVFPVPEMLAAAEALVKASHTENTYKPAVESLSRAFNLAERKADRSVQVDPDL 600

Query: 601 FENNQEKALYQAILSELTEDMHDLKGLFALSPIINDFFDNTWMTDDEHKQKNRLAIL 660
 FEN QE L+ AI L L L++FALSP+INDFFDNTWMT D+ +K NRLAIL
 Sbjct: 601 FENBQENTLFAAIQGLTLAGSAAQLBQCVFALSPVINDFFDNTWMTWAGDQALKNRLAIL 660

Query: 661 NSLVAKARTVAAPNLLATK 679
 + LV+KA+T+ APN LATK
 Sbjct: 661 SDLVSKARTIVAPNQLATK 679

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 2199

A DNA sequence (GBSx2318) was identified in *S. agalactiae* <SEQ ID 6789> which encodes the amino acid sequence <SEQ ID 6790>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2182(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD24436 GB:AF112858 MAD(P)H dehydrogenase [Bacillus
 stearothermophilus]
 Identities = 64/174 (36%), Positives = 98/174 (55%), Gaps = 6/174 (3%)

Query: 2 NTLIVNSHPDFSNPYSPTTLQEKFTIYNEHFPNHLISLNLVDCVLPETKEVLLSIIW 61
 N L + HP + S++ + + FI+ Y + P+H++ L+LY +FEI +V S W
 Sbjct: 3 NVLYITATPH- DDTQSYEMAVGKAFIDTYKQVHPDHEVTHLDLYKEYIPEIDVDFV- SGW 60

Query: 62 SKQRNGL---ELTADIEIVQAKISKDLLEQFKSHHRIVFVSEMNPNVTARAKTYIDNIFI 118
 K R G EL+ +E + +L EQF S + VFW+EM N++ K YID + +
 Sbjct: 61 GKLRSGKSPFELSDERKAKVGMNELCEQFTISADKIVFVTEMNPNFSPFVLKAYIDAVAV 120

Query: 119 AGETFKYITENGSGVGLMIDDYRLIMLESAGSIVSKQOYSYPFPPVHYLKAIFDF 172
 AG+TFKYITE G VGL+ID + L +++ G YS+G + E YL I + F
 Sbjct: 121 AGKTFKYITQQPVGLLTDK-KALHIGARQGFYSEGPAAEMGMHRYLSVINGQFF 173

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2200

A DNA sequence (GBSx2319) was identified in *S. agalactiae* <SEQ ID 6791> which encodes the amino acid sequence <SEQ ID 6792>. This protein is predicted to be glycyl-tRNA synthetase (glyQ). Analysis of this protein sequence reveals the following:

Possible site: 56

-2479-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.1364 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9521> which encodes amino acid sequence <SEQ ID 9522> was also identified.

- 10 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA005089 GB:AP001511 glycyl-tRNA synthetase (alpha subunit)
 [Bacillus halodurans]
 Identities = 222/287 (77%), Positives = 250/287 (86%)

- 15 Query: 6 LTPQEIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPTTFLRAIGPEFWNAAIVEPSRRPA 65
 + Q +ILTLQ++W+ Q C:L+QAYD ERKAGTMSPTT LR IGPEFWN AIVEPSRRPA
 Sbjct: 1 MNVQIMILTLQEIYWSKQNCILLQAYDNEKGAGTMSPTTMLRTIGPEFWNVAIVEPSRRPA 60
- 20 Query: 66 DGRYGENPRLYQHHQFQVVMKPSPSNIQELYLKSLLELGINFLEHDIRFVEDNEMENPST 125
 DGRYGENPRLYQHHQFQV+MKPS+NIQELYL SL LGINFLEHDIRFVEDNEMENPS
 Sbjct: 61 DGRYGENPRLYQHHQFQVIMKPSPTINIQELYLDLRLALGINFLEHDIRFVEDNEMENPSL 120
- 25 Query: 126 GSAGLGEVWLDGMEITQFTYFQQVGLQTGFPVTSEVYGLERLASIYQEVDSVYDIENW 185
 G AGLGEVWLDGMEITQFTYFQQVGL+ PV++E+TYGLERLASIYQ+ ++V+D+EW
 Sbjct: 121 GCMALGEVWLDGMEITQFTYFQQVGLANFVSAEITYGLERLASIYQKENVFDLENV 180
- 30 Query: 186 PGVKYGEIFTPQPEYHSKYSFESDQVMLLENFEKFSREAKRALEGLVHPAYDLYVLKCS 245
 G YG+IFTPQPEYHSKY+FE+SD ML E F +E+EA RALEE LV PAYDLYVLKCS
 Sbjct: 181 EGFTYGDIFTPQPEYHSKYTFEVSQMSALFELFSTYKEADRALLENLVFAYDLYVLKCS 240
- 30 Query: 246 HTFNLLDARGAVSVTERAGYIARINLARVAVKTFVARKKLGFPPL 292
 HTFNLLDARGA+SVTER GYI R+RNLR AK + ER+KLGFP+L
 Sbjct: 241 HTFNLLDARGA+SVTERTYIGRVNLRARKCAKTYTERKLGFPML 287

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6793> which encodes the amino acid sequence <SEQ ID 6794>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

- 40 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.2081 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 290/304 (95%), Positives = 294/304 (96%)

- 50 Query: 2 MSKKLTPQEIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPTTFLRAIGPEFWNAAIVEPS 61
 MSKKLTPQEIILTLQQ+WNDQGCMLMQAYDNEKGAGTMSPTTFLRAIGPEFWNAAIVEPS
 Sbjct: 1 MSKKLTPQEIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPTTFLRAIGPEFWNAAIVEPS 60
- 55 Query: 62 RRPADGRYGENPRLYQHHQFQVVMKPSPSNIQELYLKSLLELGINFLEHDIRFVEDNEM 121
 RRPADGRYGENPRLYQHHQFQVVMKPSPSNIQELYL SLE LGINFLEHDIRFVEDNEM
 Sbjct: 61 RRPADGRYGENPRLYQHHQFQVVMKPSPSNIQELYLSLEKLGINFLEHDIRFVEDNEM 120
- 55 Query: 122 NPSTGSAGLGEVWLDGMEITQFTYFQQVGLQTGFPVTSEVYGLERLASIYQEVDSVYD 181
 NPSTGSAGLGEVWLDGMEITQFTYFQQVGLT PVT+EVYGLERLASIYQEVDSVYD
 Sbjct: 121 NPSTGSAGLGEVWLDGMEITQFTYFQQVGLLTSPTVAEVYGLERLASIYQEVDSVYD 180
- 60 Query: 182 IEWAPGVKYGEIFTPQPEYHSKYSFESDQVMLLENFEKFSREAKRALEGLVHPAYDLY 241
 IEWAPGVKYGEIF QPEYHSKYSFESDQ MLLENFEKFE+EA RALEEGLVHPAYDLY
 Sbjct: 181 IEWAPGVKYGEIFLQPEYHSKYSFESDQVMLLENFEKFEKRALEEGLVHPAYDLY 240

-2480-

Query: 242 LKCSHTFNLLDARGAVSVTERAGYIARINRLARVVAFTFAERKKLGFPILDETRIKLL 301
LKCSHTFNLLDARGAVSVTERAGYIARINRLARVVAFTFAERKKLGFPILDE TR LL
Sbjct: 241 LKCSHTFNLLDARGAVSVTERAGYIARINRLARVVAFTFAERKKLGFPILDETRAILL 300

Query: 302 AEDD 305
AE+D
Sbjct: 301 AEDD 304

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2201

- A DNA sequence (GBSx2320) was identified in *S. agalactiae* <SEQ ID 6795> which encodes the amino acid sequence <SEQ ID 6796>. This protein is predicted to be vacB protein (vacB). Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2966 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9399> which encodes amino acid sequence <SEQ ID 9400> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE15366 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
Identities = 338/780 (43%), Positives = 485/780 (61%), Gaps = 47/780 (6%)

- Query: 4 AKAFPKLIKTIISNLESHRQL---RFDDNGSLSLQKKEAKKKEITVRGLFRANKAGPGFL- 59
A+ F +L+K + IE + R D G +K +G A+ GP FL
Sbjct: 36 AEEFKELVKALVALEKGLIVTRSDRYG-----IPEKONLIKKGKISAHAGKPAFL 87
- Query: 60 SIDQDEDMPIGKNDIAVAIDGTVAEAVVKKPADRLNGTAARVNVNIVERSLKTAVGKF 119
D D+PI N++ A+GD V + + +G+ E V+ I+ER+++ +VG +
Sbjct: 88 PEDTSLSDVIPPELNLNTAMNGDVMVRLNSQS---SGSRQSTVIRILERAIRQVVGTY 146
- Query: 120 VLDDERPKVAGYIKSNQKINQKIYIRKEPV--VLDGTBLIKVDIDKYPTRGHDYFVASV 177
+ G++ ++KI I+I K +G+++ V + YP G V
Sbjct: 145 T---ETRNPGFVTPDDKKITSDIFIPKNGKNGAAGSHKVV-VKLTSYP-EGRNNAEDEV 198
- Query: 178 RDIVGHQCDGVLDVLEVLBSMDIVSEFPEDVIAENAIIPDAPEKDLIGRVDLRQSVTF 237
I+GH+ D GID+L V+ + EFP D +A++ PD EKDL R DLR + V N T
Sbjct: 199 ETILGHKNDGIDILSVIHKGHLGGRFPADAMEQASSTPTIDEKDLKDRDLADQVIVT 258
- Query: 238 IDGADAKDLDDAVHIKLLDNHGFLGVHIAADVSYVTEGSAKNREALSRGTSVYTVDRKV 297
IDGADAKDLDDAV + LD+G +LGVHIAADV+YVTE S +++EAL ROTSvt+ DRV+
Sbjct: 259 IDGADAKDLDDAVTVTKLDGSGYIKGVHIAADVSHYVTEPIDKEALRGTSVYLVORVI 318
- Query: 298 PMLPERLNGICSLNPNLDRITQSCIMEIDQNGRVNHOITQSVINTYRMTYTVANDII 357
PM+P RLNGICSLNP +DRIT SC M I+ G+V H+I QSVI TT RMTY+ VN I+
Sbjct: 319 PMLPERLNGICSLNP+VDRLITLSCMTINSQQQVTEHRI PQSVIKTRMTYTDVNVKIL 378
- Query: 358 A-GDRERISRYKSVISVQHMVTLHHTLEAMKTRRGALNFUTSEAKIMVNDKMPVDVVI 416
DEB+ +YE +V + M L L R RGA++PD EAK+V+D+G D+VI
Sbjct: 379 VDDDEELKQYKYEPLVPMKMERLAQILRDKGDMORGAVDFDKEAKVVDDBGAVKDVVI 438
- Query: 417 RNEGIAERMIESPMLAANETVAHYARLKLFFIYRIHEEPKAEKQKPFIDYASVFGVQIQ 476
R R +AB++IE PML ANETVAEH+ + +PFYIYRIHEEP AEKIQK+ + + PG ++

-2481-

Sbjct: 439 RERSVAEKLIIEFMLVANETVAHFHMMVPIFYRIHEEPNAEKLKQKLFVTTTGYVVK 498
 Query: 477 GTATKITQSALQDFMKVQQQPGSEVLSSMILLSSMQARYSSHHNGHGLAAEYTHFTS 536
 GTA I ALQ + V+ +P +V+S ++LRS+QA+Y + GH+GL+ E+YTHFTS
 5 Sbjct: 499 GTAGNIHPRALQSLIDAVRURPBTIVISTVMAKSMKQAKYDQSLSHHGLSTFEYTHFTS 558
 Query: 537 FIRKYPDLIVHIRMIDY-DKAMDKA--DHANLPIELATQTSLSRRADAERIVEAMK 593
 FIRKYPDL+VHR+IR Y + +D+A + +A +P+IA TSS+ERRA+DAER + +K
 10 Sbjct: 559 FIRKYPDLIVHILIRTYLNGKVDRAETQKWAERLPDIAHTSSMSRRADAERETDLK 618
 Query: 594 KAEYMEETVGRFEGVAVSVKFGMFVELPWTIEGLIHVTTL-PEYTHFNEKITLQGEK 652
 KAEYM + +GEFP+G+++SV FGMFVELPWTIEGL+HV+ + +Y F+B+ + GE+
 Sbjct: 619 KAEYMLDKIGEEFGMISSVTFGMFVELPWTIEGLIHVSPMTDDYFIRBQHFAMIGER 678
 15 Query: 653 SGKVFVVGQKIKLLRSDEKETGLIDFYLEPSDFDIVEKVSRSRGRHNRSSKREHQR 712
 +G VFR+G +I VK++ +K+ +IDF+ + +G P R + +
 Sbjct: 679 TGVFRIGDEITVKVVDVNRDERNIDFEV-----GMKGTFRPRFELD--- 721
 20 Query: 713 ISDRDNKNKNTSKKFKARPKRNSDSKSHRHKDDRTGTGSKKTKPKFYKVAKGQKRR 772
 S R K ++K+ S + S K + T EKK K+ F +K +K+K
 Sbjct: 722 -SSRSRKRGGKPKRKRQSTNPTVSPAPS-EEGGEWFTPKKKKKRGRQAKRQKRRKKK 779

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6797> which encodes the amino acid sequence <SEQ ID 6798>. Analysis of this protein sequence reveals the following:

25 Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.0811(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 579/773 (74%), Positives = 664/773 (84%), Gaps = 22/773 (2%)
 35 Query: 1 MAGAKAPFKLIKTISNLSHRQLRFDWNGSLQKKEARKKEITVGLFRANKAGPGLS 60
 MAGAK FP LIKTIS +BS LRF D+GSL+L+K+ KKE IV+G+FRANKAGPGL
 Sbjct: 27 MAGAKHPFLIKTISKMSQSLRFSDDGSLARKERHKKEPTVGQVFRANKAGPGLH 86
 40 Query: 61 IDQEDDDMFIGNDIAYIDSDTVEANVKKPADRLNGTAARVAVIVERSIKELVCKFV 120
 +D+EDDMFIC+ND+ YAIIDGIVE VVKKPADRL GTAAEA+VV IV+RSIKT VG F+
 Sbjct: 87 VDEMDDMFICRMDVGYAIDGTVFVVVKKPADRLGTAAEA+VAVIVDRSIKTAVVTFI 146
 45 Query: 121 LDDERPKYAGYIKSKQKINOKYIIRKEPVFLDGTIELIKVDIDKYPIRGHDYFVASVRDI 180
 LDD++PKYAGYI+SKQKI QKIYI+KEPVVL GTIELIKVDIDKYP RGHDFVASVRDI
 Sbjct: 147 LDDDKPKYAGYIRSKQKIQKIYI+KEPVVLGTIELIKVDIDKYPIRGHDYFVASVRDI 206
 50 Query: 181 VHQGDVGIDVLEVLSSMDIVSEFPEDVIAKANAIAPDPTKDLIGRVDLRQGVPTTID 240
 VHQGDVGIDVLEVLSSMDIVSEFP +V+AEANAI +APT KDLIGRVDLRGE T TID
 Sbjct: 207 VHQGDVGIDVLEVLSSMDIVSEFPAAVLAEANAISEAPTAADLIGRVDLRGETTTID 266
 55 Query: 241 ADAKDLDLDAVHIKLLDNGHVELGVHIADVSYYVTKSALNRALSRCSTSVYVDRVFMIL 300
 ADAKDLDLDA+HIKLLDNG+ELGVHIADVSYYVTKSAL++EA++RCTSVYVDRVFMIL
 Sbjct: 267 ADAKDLDLDAIHIKLLDNGHVELGVHIADVSYYVTKSALDKAELANGSTSVYVDRVFMIL 326
 60 Query: 301 PERLSNGICSLNPNLDRLTQSCIMEIDQGRVNVQITQGVINTTYHMTTAVNDIAGD 360
 PERLSNGICSLNPN+DRLTQS +MEI+ G VNV+QI QSVI TTYHMTT+ VID+IAGD
 Sbjct: 327 PERLSNGICSLNPNIDRLTQSLMIRNSQGVNVNVIQIQSVIKITTYHMTTSTVNDIAGD 386
 65 Query: 361 EETCSYSEIVSSVQRMVHTLHTLEAMTRRGALNPDTSBAKIMVNDKGMVDIVRNRG 420
 EE B+ SI V MV LH LRAMR++RGALNPDT BAKI+VNDKGMVD+V+R RG
 Sbjct: 387 EEALQEPASIADDVTLMVHLHTLEAMRSKRGALNPDTPQAKIIVNDKGMVDIVRNRG 446
 Query: 421 IARMIESVFLAANETVAHFYARLKLPIFYRIHEEPKAEKQKPFIDYASVFGVQIGSTAT 480
 IARMIESVFLAANE VAER+R+ KLPIFYRIHREPKAEKQK+FDIYAS FG+ IQGTA

-2482-

Sbjct: 447 I A E R M I E S F M L A A N E C V A S H F A K A K L P F I Y R I H E S P K A S K L Q Q F D Y A S T P G I H I Q S T A N 506
 Query: 481 K I T Q S A I Q D F M K I V Q Q Q P Q S E V L S M L L R S M C Q A R Y S E I N H I G Y L A A R Y T T H T S P I R R 540
 K I + Q A L Q F M K V + Q Q P + E V L + M M L L R S M C Q A R Y S E I N H I G Y L A A R Y T T H T S P I R R 540
 5 Sbjct: 507 K I S Q S A I Q A F M A K V E G Q P G A E V L N M M L L R S M C Q A R Y S E I N H I G Y L A A R Y T T H T S P I R R 566
 Query: 541 Y P D L L V H R M I R D Y D D K A M D K A D H F A N L I P E A T I Q T S S L E R R A I D A E R I V E A M K K A E Y N M E E 600
 Y P D L L V H R M + R + Y + + + K D H F A + I P E + N I + S L E R R A I D A E R + V E A M K K A E Y N E 600
 10 Sbjct: 567 Y P D L L V H R M V R E Y N Q P S Q E K R D H F A Q I I P E L A T S S S Q L E R R A I D A E R V E A M K K A E Y N A E 626
 Query: 601 Y V G S E P G S V A S V V K F G M F V E L P W T I E G L I H V T T L P E Y T H N E R T L I L Q Q E K S G K V F R V G 660
 Y V G S E F + G + V + S V V K F G F V E L P W T I E G L + H + T + L P E Y T H N E R T L + I L Q Q E K S G K V F + V G 660
 Sbjct: 627 Y V G S E F D G I V S V V K F G F F V E L P W T I E G L V H I T L S L P E T T H N E R T L S L Q Q E K S G K V F V G 686
 15 Query: 661 Q Q I K V K L I R S D K E T G D I D F T Y I L P S D F D I V S K V S K S R E G R P N R S K R E H Q I R I S D R N K M 720
 Q I + V K L + + + D K E T S D I D F + Y L P S D F D + V E K + S + R + R K + 720
 Sbjct: 687 Q P I R V K L V K A D K E T G D I D F E I L P S D F D V E K I R M S D K A S R A D R - - - - - R K S 732
 Query: 721 I N T S K K K A S R K P K R N S D K S H H H K D D R T T G S T K K T K K P F K V K V A K K G Q R K S 773
 + S K + + E K + + K T G T K E + K K F P K A K K + + + S 773
 20 Sbjct: 733 S K S K G T K K E K P E V A K A K - - - - - T G K T K E G S K K P F K E Q A K K S R K S 777

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2202

A DNA sequence (GBSx2321) was identified in *S. agalactiae* <SEQ ID 6799> which encodes the amino acid sequence <SEQ ID 6800>. This protein is predicted to be VacB homolog (smpB). Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 30
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2988 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23745 GB:AF052209 VacB homolog [Streptococcus pneumoniae]
 Identities = 121/155 (78%), Positives = 139/155 (89%)
 40 Query: 1 M V K Q G N V V A Q N K A H D Y T I V E T I E A G I V L T G T E I K S V R A A R I T L K D G Y A Q I K N G E A W L 60
 M K G + G V V A Q N K A H D Y T I V + T + E A G + V L T G T E I K S V R A A R I L E D G + A Q + K N G E W L 60
 Sbjct: 1 M A K S E G K V V A Q N K A R H D Y T I V D T L E A G M V L T G T E I K S V R A A R I N L K D G F A Q V K N G E W L 60
 45 Query: 61 I N V H I T P Y D Q G N I W N Q D P D R T R K L L A K R E I K L S N E L K G T G M T I V L P L K V Y L K D G F A K V L 120
 N V H I P Y + G N I W N Q + P + R R K L L K + + I + K + E K S T G M T I V L P L K V Y + K D G + A K + L 120
 Sbjct: 61 S N V H I A P Y E B G N I W N Q E P E R R R K L L A H K Q I Q K L Q E T K G T G M T I V L P L K V Y I R D G Y A K L L 120
 Query: 121 L G L A K G K H D Y D K R E S I K R R E Q N R D I A R L K N Y N S R 155
 L G L A K G K H D Y D K R E S I K R R E Q N R D I A R + K N R 155
 50 Sbjct: 121 L G L A K G K H D Y D K R E S I K R R E Q N R D I A R V M K A V N Q R 155

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6801> which encodes the amino acid sequence <SEQ ID 6802>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 55
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2918 (Affirmative) < succ>

-2483-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 124/155 (80%), Positives = 145/155 (93%)

Query: 1 MVEGQGNVVAQNKKAHHDTTIVETIEAGIVLTQYTIKSVRAARITLKGQYAIKNGEAML 60
 M KG+G+++AQNKKA HDY IVET+EAGIVLTQYTIKSVRAARI LKDG+AQIKNGEAML
 10 Sbjct: 1 NAKGEHILACNKKARHDYITVETIEAGIVLTQYTIKSVRAARITLKGQYAIKNGEAML 60

Query: 61 INVHTPYDQGNINWQDPQTRKILLKKREIEKISNELKGTGMLVPLKVLYLKDGFPAVL 120
 +NVIH P++QGNINW DP+RTRKILLKKREI ++NELKG+GHTLVPLKVLYLKDGFPAVL
 10 Sbjct: 61 VNVHIAPEQGNINWADPERTKILLKKREITHLANELKSGMTLVPLKVLYLKDGFPAVL 120

15 Query: 121 IGLAKGKHEYDKRESIKRREQNRIARQLKNYNSR 155
 +GLAKGKH+YDKRE+IKR+Q RDI +Q+K+YN+R
 Sbjct: 121 IGLAKGKHEYDKRESIKRREQNRIARQLKNYNSR 155

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 20 vaccines or diagnostics.

Example 2203

A DNA sequence (GBSx2322) was identified in *S. agalactiae* <SEQ ID 6803> which encodes the amino acid sequence <SEQ ID 6804>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.6876 (Affirmative) < succ>
 30 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 35 vaccines or diagnostics.

Example 2204

A DNA sequence (GBSx2323) was identified in *S. agalactiae* <SEQ ID 6805> which encodes the amino acid sequence <SEQ ID 6806>. This protein is predicted to be d-serine/d-alanine/glycine transporter (cycA). Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have a cleavable N-term signal seq.

40	INTEGRAL	Likelihood = -9.02	Transmembrane	71 - 87 (62 - 90)
	INTEGRAL	Likelihood = -8.92	Transmembrane	320 - 336 (316 - 344)
	INTEGRAL	Likelihood = -8.33	Transmembrane	254 - 270 (251 - 275)
45	INTEGRAL	Likelihood = -6.00	Transmembrane	158 - 174 (154 - 175)
	INTEGRAL	Likelihood = -2.76	Transmembrane	197 - 213 (196 - 213)
	INTEGRAL	Likelihood = -2.50	Transmembrane	117 - 133 (116 - 136)
	INTEGRAL	Likelihood = -1.38	Transmembrane	282 - 298 (279 - 298)
50	INTEGRAL	Likelihood = -0.32	Transmembrane	342 - 358 (342 - 360)

----- Final Results -----
 bacterial membrane --- Certainty=0.4609 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-2484-

A related GBS nucleic acid sequence <SEQ ID 9397> which encodes amino acid sequence <SEQ ID 9398> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB14651 GB:Z99117 amino acid permease [Bacillus subtilis]
 Identities = 165/361 (45%), Positives = 227/361 (62%), Gaps = 17/361 (4%)

Query: 1 MGIFLT-LSYNISLIFIGMABITAVGEYVQFWFFWPNWSNIIQIVFLAILSSINLIAVKAF 59
 M F+T +YW I +MA++TAVG Y Q+H P+ P W+ ++ L IL +NL VK F

10 Sbjct: 95 MAAPITGWYWFPCWISLAMDITAVGIYQWLEDVQWNLGLLAILLIMHATVKLF 154

Query: 60 GTEFWFAMIKVIAILGLIATGIPMWLINFUTGHGYHASISNITHFWFPGKGLNFPMA 119
 GE EFWFA+IKVIAIL LI TGI ++ F G AS++N+ +H FF G F ++

15 Sbjct: 155 GELEFWFALIKVIAILALIVTGILLIANGPSAASG- PASLNNLSHGQMFPGMHOFILS 213

Query: 120 FQWVFAYLAIEFVGVTSETANPRKVLKAIQEIPIRIILFYAGSLIAIMAFWQQLF 179
 FQMV FA++ IE VG+T ET NP+KV+PKAI +IP+RI+LEF G+L IM I+PW L

20 Sbjct: 214 FQWVFVAFVGLIEGLIENGETNPQKVLKAIQIPVRIILFYVGLFVIMCIPKAVLN 273

Query: 180 VNESFFVTVFKLAGIKWAAALINFVVLTSAAASAINSTLYSTGRHLFQLANE--SPNALTK 237
 NESFFV VF GI AA+LINFVVLTSAAAS NS L+ST R ++ LA + +P L K

25 Sbjct: 274 VNESFFVQVSAVGIVAAALINFVVLTSAAASAINSTLSERMVYSIAKHHPATGLKK 333

Query: 238 ALKLDQLSRQVFSRAIILS--AVIVGASALISVLPGISDAFSLITASSGGVYISIVLI 295
 L+ +VPS R+ S R++G S L ++P F+LIT+ S+ +I L+ +

30 Sbjct: 334 -----LTSSNVPSNALFFSSAILLIGVS-LNYLMP--EQVFTLITSVSTICPIFWGIT 384

Query: 296 MIAHWKVRKS--PDFMEDGYKMPYKILSPITLLFPLFVFSVSLFQDSTYIGAIGATTWII 354
 +I H KYRK+ + + +IMP Y + + +TL F F+ V L L + T I +W +

Sbjct: 385 VICHLYKRYKTRQHEAKANKPMFFYPLSNYLTALFLAFLILVLALANDTRIALFVTVFVWF 445

There is also homology to SEQ ID 4070:

Identities = 286/364 (78%), Positives = 322/364 (87%), Gaps = 1/364 (0%)

35 Query: 2 GIPFLT-SYNISLIFIGMABITAVGEYVQFWFFWPNWSNIIQIVFLAILSSINLIAVKAFGE 61
 G F LSYNISLIFIGMABITAVG YVQFWFP WP+N+IQ+VFL +LSSINLIAV+ FGE

Sbjct: 101 GYPGCLSYNISLIFIGMABITAVGAYVQFWFSWFWALIQVLVLVLS INLIAVRVFG 160

40 Query: 62 TEFWAMIKVIAILGLIATGIPMWLINFUTGHGYHASISNITHFWFPGKGLNFPMA 121
 TEFWAMIK+AIL IAT IPMWLT F+T H HAS+SHI +HF FF GKL FFWAQ

Sbjct: 161 TEFWAMIKILAILALIALIATAIIPWLTGFET- HGHASLSNIFDHFVWFPGKGLKFFFAQ 219

Query: 122 MVFPAYLAIEFVGVTSETANPRKVLKAIQEIPIRIILFYAGSLIAIMAFWQQLFVN 181
 MVFPAY AIEFVG+TSETANPRKVLKAIQEIPI RI++FY G+L+IMAI PW QLFV+

45 Sbjct: 220 MVFPAYQAIIEFVGITSETANPRKVLKAIQEIPIRIIVFYVGLVNSINIAVPMIQLFVD 279

Query: 182 ESPPFVTVFKLAGIKWAAALINFVVLTSAAASAINSTLYSTGRHLFQLANESPNALTKALK 241
 ESPPV VKL GIKWAAALINFVVLTSAAASAINSTLYSTGRHL+Q+ANE+PNAL LK+

50 Sbjct: 280 ESPPVMVKLIGIKWAAALINFVVLTSAAASAINSTLYSTGRHLYLQIANETPNALNKLKI 339

Query: 242 DQLSRQVFSRAIILSASAVIVGASALISVLPGISDAFSLITASSGGVYISIVLIMIAHWK 301
 + LSRQ VFSRAIILSASAV+VG SALI++LEG+DAFSLITASSGGVYI+Y L MIAHWK

Sbjct: 340 NTLSRQVFSRAIILSASAVVGVISALINILPGVADAFSLITASSGGVYIATALTMIHWK 399

55 Query: 302 YRKSPDFMEDGYKMPYKILSPITLLFPLFVFSVSLFQDSTYIGAIGATTWIIIGFGLYSH 361
 YR+S DFM DGY MP YK+ +P+TL FP FVF+SLFLQ+STYIGAIGATTWII FG+YS+

Sbjct: 400 YRKSKWMDGYKMPYKIVTPTLAFAPFVPSLFLQESTYIGAIGATTWIIIPGIYGN 459

Query: 362 FKHK 365
 K K

60 Sbjct: 460 VKFK 463

-2485-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2205

A DNA sequence (GBSx2324) was identified in *S. agalactiae* <SEQ ID 6807> which encodes the amino acid sequence <SEQ ID 6808>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.33	Transmembrane	194 - 210 (191 - 215)
INTEGRAL	Likelihood = -5.47	Transmembrane	17 - 33 (14 - 38)
INTEGRAL	Likelihood = -5.15	Transmembrane	125 - 141 (119 - 144)
INTEGRAL	Likelihood = -3.88	Transmembrane	155 - 171 (153 - 176)
INTEGRAL	Likelihood = -1.38	Transmembrane	95 - 112 (94 - 114)
INTEGRAL	Likelihood = -0.43	Transmembrane	49 - 65 (49 - 65)

----- Final Results -----

bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]

Identities = 80/214 (37%), Positives = 122/214 (56%), Gaps = 3/214 (1%)

Query: 4 FFSNIRTEIPQMFLLIHSLLSLVFLMMLTLVNRDKPLYKTINSILLGLQLITITYVFP 63
FF+ T+ E+ L + + ++L + R+K +Y+ + IL +QLI +Y W++
Sbjct: 7 FFTQTATKFFKFDLFWVSLFTLIALTFYTAHRYREKKVYQRFQHLQTLQLILKGYW 66

Query: 64 WAKLPLESSEPLHYCRIGMFVLLARPGI--LKDYFALLGVVGVLMIHDPDFPYQPLH 121
+PLSESLP YHCR+ MEVULL PQ K YFALLG G + A ++P Y F H
Sbjct: 67 WEMPLSESLFPYHCRMAPFVLL--PQSKYQYFALLSTPGTLAARVYVPVDAIPPH 125

Query: 122 VTNIFFFIHGFALFVLSLHMTQSNLDKINPKLIQLTLLNMSLIPINLTGGNYGFM 181
+T + F GH AL SL++L Q N L+ K I +T +N + +NL+TGG+YGF+
Sbjct: 126 ITILSFPGHLLALLGSLVLLVLRQYNARLLDVKGFLMTFALNALIPVNLVTGGDYGFL 185

Query: 182 MKTFPLGITNPFMLFVITVLLSFVLLPVKQIQP 215
K P++G N +V+ +L + K+I +
Sbjct: 186 TKPFLVGDHGLVANYLLVSVLVAVATISLTKILE 219

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6809> which encodes the amino acid sequence <SEQ ID 6810>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.25	Transmembrane	16 - 32 (11 - 39)
INTEGRAL	Likelihood = -3.45	Transmembrane	154 - 170 (153 - 173)
INTEGRAL	Likelihood = -3.08	Transmembrane	96 - 112 (94 - 112)
INTEGRAL	Likelihood = -1.91	Transmembrane	191 - 207 (191 - 209)
INTEGRAL	Likelihood = -1.12	Transmembrane	71 - 87 (71 - 87)

----- Final Results -----

bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]

Identities = 90/231 (38%), Positives = 128/231 (54%), Gaps = 7/231 (3%)

Query: 3 FFAIDPIGLPHTSLIPYLSLLIALLNLVLTQAYRK-S-HRYVFLPLQLSQVIGLYTWY 61

-2486-

FF P L ++S L L L P T YR K ++ PF LQ Q+I LY WY
 Sbjct: 7 FFTTQATKPKPKFDLFWYVS-LFTLLAITYFYAHRYREKKVYQRFFQILQTVQLLLYGMV 65
 Query: 62 VLRGFPDLDEALFLYHCRIMLAIFFLPDRNKFKQLRNVLIGIGTFLALL--SPDLFPRL 119
 + PL E+LP YHCR+AM + LP ++K+EQ F +IG GT A + PD YFF
 Sbjct: 66 WWHMPLSESLPFFYHCRIMAFVLLLEGQSKYQYFALLGTFTGLAAPVYFVPDAPYFP- 124
 Query: 120 WHVANVSFYGHYALLVNGLIYLLRFYDASQLRLSVVRYLATVNFLLLVSLATGNGY 179
 H+ +SF FGH ALL N L+YLLR Y+A L + + +N L+ +V+L T G+YG
 Sbjct: 125 -HITLISFTIGHLAALLNSLYLLRQYNARILLVKGIFLMTFALNALIFVNLVYGDYG 183
 Query: 180 FVMDIPVHTRHLLNFVITVSGLTFMVKITEYFYLKPGEAQQLALAFSKE 230
 P+ P++ L+ N+++V+ L + +T+ L+F AQ+ KE
 Sbjct: 184 FLTKPFLVGDHGLVANYLLVSIIVATISLTKKI-LEFFLAQEAERKIVKE 233

An alignment of the GAS and GBS proteins is shown below.

Identities = 70/216 (32%), Positives = 117/216 (53%), Gaps = 1/216 (0%)

Query: 2 IEFFSNIRTEIPQMPILLIHSLLSFLPFLMWLTVNRDKPLYKTIINSILLGLQITITYN 61
 ++FF+ +P L+ + L+ L++LT ++ + L Q+I +YTN
 Sbjct: 1 NDFPAIDPIGLPHTSLIFLYSSLLIALLVFLTPQAYRLKSHRYFFFLQLSQVIGLYTN 60
 Query: 62 FFWACFLSESLFLYHCRIMFVVL-LARPGILKDYFALLGVVGGVLAMHDFPIYQFL 120
 + FL E+LFLYHCRIM + L K F +L+ G L++ ED YP++
 Sbjct: 61 YVLRGFPDLDEALFLYHCRIMLAIFFLPDRNKFKQLRNVLIGIGTFLALLSPDLFPRLW 120
 Query: 121 HVTNIFPFIHGHALFVLSLHMTQSNLDKLNPKLIQITLNLINSLIPINLLTGNNGYF 180
 HV N+ P+ GH+AL V L++L+ + +L +++ +N L+ ++L T GNYGF
 Sbjct: 121 HVANVSFYGHYALLVNGLIYLLRFYDASQLRLSVVRYLATVNFLLLVSLATGNGYF 180
 Query: 181 NMKTFILGITNPFNLIFVITLSSFLVFLVKQIPK 216
 +M P++ + LN IVT+ L+F+V + +K
 Sbjct: 181 VMDIPVHTRHLLNFVITVSGLTFMVKITEYFYLK 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2206

A DNA sequence (GBSx2325) was identified in *S. agalactiae* <SEQ ID 6811> which encodes the amino acid sequence <SEQ ID 6812>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3297 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2207

A DNA sequence (GBSx2326) was identified in *S. agalactiae* <SEQ ID 6813> which encodes the amino acid sequence <SEQ ID 6814>. This protein is predicted to be oxalate:formate antiporter (oxIT-2). Analysis of this protein sequence reveals the following:

-2487-

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

5	INTEGRAL	Likelihood = -7.80	Transmembrane	380 - 396 (376 - 399)
	INTEGRAL	Likelihood = -7.43	Transmembrane	291 - 307 (284 - 310)
	INTEGRAL	Likelihood = -5.63	Transmembrane	169 - 185 (163 - 186)
	INTEGRAL	Likelihood = -4.99	Transmembrane	226 - 242 (223 - 245)
	INTEGRAL	Likelihood = -4.19	Transmembrane	46 - 62 (39 - 63)
10	INTEGRAL	Likelihood = -4.09	Transmembrane	311 - 327 (308 - 329)
	INTEGRAL	Likelihood = -1.49	Transmembrane	261 - 277 (260 - 278)
	INTEGRAL	Likelihood = -1.06	Transmembrane	133 - 149 (133 - 150)
	INTEGRAL	Likelihood = -0.85	Transmembrane	98 - 114 (98 - 114)
	INTEGRAL	Likelihood = -0.06	Transmembrane	77 - 93 (77 - 93)

----- Final Results -----

15	bacterial membrane --- Certainty=0.4121 (Affirmative) < succ>
	bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20	>GP:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus lactis]		
	Identities = 220/398 (55%), Positives = 306/398 (76%), Gaps = 3/398 (0%)		
25	Query: 5	NRYVAVSGVVLHLMIGSTYANGSVFRPIISETGNDISSVFASFSLAIFCLMSAAPMGH 64	
	Sbjct: 4	NRYVVA +GV+ HLM+GS YANGSVF NPI + GN SSV+ AFS+AI+ LQMSAAPMG 63	
30	Query: 65	LVERFGPRINGMISAILYGAGNVLTGLAIETQQLMLLVAYGILGIGLSSGYITPVSTI 124	
	Sbjct: 64	VVEKIGPRLTGTIASFLYGTGTMTGWAHQNSINLLYLSYGVIGGLGAGYVTPVSTI 123	
35	Query: 125	IKWFPDRRLGATGPAIMGPGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFPVMHIAQSF 184	
	Sbjct: 124	IKWFPDKRLGATGLAIMGPGFAALITGPVQLMAGVGLSCTFYILGTFFVFMILAAQF 183	
40	Query: 185	IKQPPQEKITILTHQKIONMNSQIITG--LKANAIRKSTFYIILWTLFINISCGGLI 242	
	Sbjct: 184	IVR-PHIALSSTTENSISQKQTRITRGPFLTANQALTKSPFLMIMFFINTCGILV 242	
45	Query: 243	SAASPMAGDLGYSABSAALLVGVLGIFNGPQRLNWSLDYIGRPILFIFILVNFMT 302	
	Sbjct: 243	SAASPMAG + G S ++AA++VG++G+FMGPRL+WA+LSDYIGRP TF +PI++ +M 302	
50	Query: 303	SSLFLSFNAIVFAIMSIIMTCYAGGFSLLPAYLSDIPGTELATLHGYSLTAWAAGLF 362	
	Sbjct: 303	SAILIFKLPLFLVIALCLMSCYGAGFSVIPAYLGDVPGTELGAAGVGYLTAWAAGV 362	
55	Query: 363	GPLLLSKTYSWGSYQLTLMVFGFLFLGLLLSLYLRK 400	
	Sbjct: 363	GPLLLS T+ ++Y LTL F + L LL+S +++ 400	

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6815> which encodes the amino acid sequence <SEQ ID 6816>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

55	INTEGRAL	Likelihood = -12.95	Transmembrane	289 - 305 (282 - 321)
	INTEGRAL	Likelihood = -11.83	Transmembrane	376 - 392 (372 - 397)
	INTEGRAL	Likelihood = -8.55	Transmembrane	163 - 179 (160 - 189)
	INTEGRAL	Likelihood = -7.75	Transmembrane	227 - 243 (221 - 247)
	INTEGRAL	Likelihood = -5.89	Transmembrane	44 - 60 (41 - 67)
60	INTEGRAL	Likelihood = -1.38	Transmembrane	310 - 326 (309 - 327)
	INTEGRAL	Likelihood = -0.90	Transmembrane	353 - 369 (353 - 369)
	INTEGRAL	Likelihood = -0.37	Transmembrane	138 - 154 (138 - 154)
	INTEGRAL	Likelihood = -0.06	Transmembrane	98 - 114 (98 - 114)
	INTEGRAL	Likelihood = -0.00	Transmembrane	259 - 275 (259 - 275)

-2488-

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.6180 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- >GP:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus
 lactis]
 10 Identities = 222/399 (55%), Positives = 305/399 (75%), Gaps = 3/399 (0%)
 Query: 3 KIRKYLIATAGILHLMLGSTYANVSVRNPILQETGMDQAPVAFAPSAIFCLGLSAAFM 62
 KT RY++A AG++ HLM+GS YANSV+ NPI ++ GN ++ VA APS+AI+ LG+SAAPM
 Sbjct: 2 KIRYVYVAFAGVFMHLMIGSVYANSVFTNPITAKQNGWBSSEVALAFSAITVFLGMSAAM 61
 15 Query: 63 GNLVEQYGPRLTGTVSAILYASGNMHTGLAIDRKKIWLILYIGYGVIGSLGLGNGYITPIS 122
 G +VE+ GPRLTGT+++ LY +G ++TG AI + IWLLY+ YGVIGSLGLGNGY+TP+S
 Sbjct: 62 GRVVEKIGPRLTGTIASFLYGTGTITGWAHQNISWLLYLSYGVIGSLGLGNGYVTPVS 121
 20 Query: 123 TIIRKFPDKRGMAITGPAINGPFGASLLTSPICQNLITETBGLVATFYLLGLIYLIVMLFAS 182
 TIIRKFPDKRG+ATG AINGPFGPA++LT P+AQ L+ + GL TFLYLG Y ++ML A+
 Sbjct: 122 TIIRKFPDKRLATGLAINGPFGAAMLTPGPAQLMASVGLGCTFYLLGTYFFVIMLLAA 181
 25 Query: 183 QLIKPTAABIAILDKKRLQ--NNSYLIBG--MTAKEALKTRSFYFCMLVILFINITOCGLG 239
 Q I++P A + + Q + L G +TA +ALKTRSF LN++ FINITOG+GL
 Sbjct: 182 QPIVRPNIALSSITENSISQKKGTRLTGPGELTANQALKTNSPTFLIMINFINITOCIGL 241
 Query: 240 ISVVPMAQDLTGMSFENSAIVVGAMGIFNGPGRLWASLSDYIGRRVTIVLLFLVSIIM 299
 +S +PMAQ +TGMS +AI+VG +G+FNPGRL+WA+LSDYIGR T +P++ I+M
 30 Sbjct: 242 VEASAPMAQSGMTGMSVQTAAIMVGIIGLFGNPGRLIWAITSYDYGRRPATSAIFLIDVIM 301
 Query: 300 TISLIFAHSSILFIMISATIMTCYAGGFSLLPFLYSLDLPKAKELATLHGYLITAWAIAAL 359
 +++ L+P+I++ LN+CYGAGFS+IP YL D+FG KKL +HGY+ITAWA A +
 35 Sbjct: 302 LSAILIFKLELFLVIALCLLMSYCGAGFSIPAYLSDIPGFTKELGAVHGVLITAWAAGV 361
 Query: 360 TGPMLLSITVETHNYLLTLCVPFVLYILGLMVALRKK 398
 GP+LLS+T + HNY LTL FI++ +L L+++ +++
 Sbjct: 362 VGPELLSLTHQLFNHYTLTAAFILIDLLALLISFWIQR 400

- 40 An alignment of the GAS and GBS proteins is shown below.

- Identities = 252/400 (63%), Positives = 329/400 (82%), Gaps = 2/400 (0%)
 Query: 1 MKNLNRYVAVSGVVLHMLGSTYANVSVRNPISSETGMDISSVPAPSAIFCLGMSAA 60
 M+ RY++A +G++LHMLGSTYANSV+RNPI+ ETGMD + V+APSAIFCLG+SA
 45 Sbjct: 1 MEKTKRYLIATAGILHLMLGSTYANVSVRNPILQETGMDQAPVAFAPSAIFCLGLSAA 60
 Query: 61 FMGHLVERPQPRIMGISAILYAGGNVLTGLALETQQLMVLVYAGVIGSLGSSGYITP 120
 FMG+LVE++GPR+ G +SAILY +GN+LGTGLAI+ +++WLLY+ YG++GG+GLG+GYITP
 Sbjct: 61 FMGNLVEYGPRLTGTVSAILYASGNMHTGLAIDRKKIWLILYIGYGVIGSLGAGGYITP 120
 50 Query: 121 VS+TIIRKFPDKRLATGPAIMGFGFASLVTSPICQNLIRIGVGKTFYILGLVYFFVMMI 180
 +STIIRKFPD+RG+ATGPAIMGFGFASL+TSP+AQ L+ +G+ TFI+GLI+Y VM+
 Sbjct: 121 IS+TIIRKFPDKRGMAITGPAIMGFGFASLLTSPICQNLITETBGLVATFYLLGLIYLIVMLF 180
 55 Query: 181 ASQFIKQPQKEITILTHDGKKNMANSQIITGLKANAIAIKSTFYIIMTLFINISCGLG 240
 ASQ I +P +I IL D K+ NS +I G+ A A+K+K+FY +N+ LFINI+CGLG
 Sbjct: 181 ASQLIKPTAABIAIL--DKKKLNNSYILIRGMTAKEALKTRSFYCLMVLFINITOCGLG 238
 Query: 241 LISAASPMADLIGYSABSAALLVGVLSIFNGPGRGLWASLSDYIGRFLTFIIPVNPFI 300
 LIS +PMAQL G S E +AI+VG +GIFNGPGRGL+WASLSDYIGR +T I+L+V+ I
 60 Sbjct: 239 LISVVPMAQDLTGMSFENSAIVVGAMGIFNGPGRLWASLSDYIGRRVTIVLLFLVSIIM 298
 Query: 301 NTSSLPLSPNAIVEFAIAMSILMTCYAGGFSLLPAYLSDIPGKELATLHGYSILTAWAIA 360
 MT SL + ++++F I+++ LMTCYAGGFSL+P YLSD+PG KELATLHG YLITAWAIA
 65 Sbjct: 299 MTISLIFAHSSILFIMISATIMTCYAGGFSLLPFLYSLDLPKAKELATLHGYLITAWAIA 358

-2489-

Query: 361 LFGPILLSKITYSWNSYQLTLMVVGFLFLPGLLLSLYLRK 400
 L GP+LLS T W +Y LTL VP L++ GL+LL L+K
 Sbjct: 359 LTGPMLLSITVWHBYLYLTLCVFTVLTYLGLMVAJLRK 398

- 5 A related GBS gene <SEQ ID 8995> and protein <SEQ ID 8996> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5
 Mcg: Discrim Score: 5.06
 GVH: Signal Score (-7.5): 4.38
 Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 10 value: -7.80 threshold: 0.0
 INTEGRAL Likelihood = -7.80 Transmembrane 380 - 396 (376 - 399)
 INTEGRAL Likelihood = -7.43 Transmembrane 291 - 307 (284 - 310)
 15 INTEGRAL Likelihood = -5.63 Transmembrane 169 - 185 (163 - 186)
 INTEGRAL Likelihood = -4.99 Transmembrane 226 - 242 (223 - 245)
 INTEGRAL Likelihood = -4.19 Transmembrane 46 - 62 (39 - 63)
 INTEGRAL Likelihood = -4.09 Transmembrane 311 - 327 (308 - 329)
 INTEGRAL Likelihood = -1.49 Transmembrane 261 - 277 (260 - 278)
 20 INTEGRAL Likelihood = -1.06 Transmembrane 133 - 149 (133 - 150)
 INTEGRAL Likelihood = -0.85 Transmembrane 98 - 114 (98 - 114)
 INTEGRAL Likelihood = -0.06 Transmembrane 77 - 93 (77 - 93)
 PERIPHERAL Likelihood = 0.42 352
 modified ALOM score: 2.06
 25 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 ORF02272(313 - 1500 of 1818)
 GP_7107009[gb][AAP36228.1][AP168363_4][AP168363(4 - 400 of 421) oxalate:formate antiporter
 [Lactococcus lactis]
 %Match = 38.5
 %Identity = 55.4 %Similarity = 79.1
 Matches = 220 Mismatches = 81 Conservative Sub.s = 94
 40
 216 246 276 306 336 366 396 426
 GK*IC*AEW*TIQFFDNIFITNYIFINK*VRF*RDCLKNRYVAVSGVVLHMLGSTYAWSVFRPIISETGWDIS
 ||||| :||:||||| ||||| ||| :|||
 MKTNRYVVAFAGVMPFLMIGSVYANRVPTNPLAKQNGWSES
 45 10 20 30 40
 456 486 516 546 576 606 636 666
 SVSFAFSLAIFCLGMSAAFHGLVERFPRIMKISAILYGMGNVLTLGAIETQQLMLLYVAYGILGGIGLGSYITPV
 ||:||||:||||| :||:||||| :||: ||| :||| || :|||:||||| :||: |||
 50 SVLAIFSIALYFLGMSAAFVKVEKIGERLITGISFLYTGITGIMTGAITHQNSIMLLYLSGVIGIGLGLGAYVTVPS
 60 70 80 90 100 110 120
 696 726 756 786 816 846 876 906
 TTIKNWFDPRKGLATGPAIMGFGFASLTVSPLAQSLMIRIGVGKTFYILGLVYFPVMIMASQFKOPPOSKITILTHDGKK
 ||:||||:||||| :||:||||| :||: ||| :||: ||| :||: ||| :||: ||| :||: ||| :||: ||| :||: |||
 55 TTIKNWFDPRKGLATGPAIMGFGFASLTVSPLAQSLMIRIGVGKTFYILGLVYFPVMIMASQFKOPPOSKITILTHDGKK
 140 150 160 170 180 190 200
 936 960 990 1020 1050 1080 1110 1140
 NMANSSQITG--LKNAAIKSKTFYIIMLTFLNISCGLGISAASPMADLAGYSARSAAALLVGVGLIIFNGPRLLIAS
 :||: ||| :||: ||| :||: ||| :||: ||| :||: ||| :||: ||| :||: ||| :||: ||| :||: ||| :||: |||
 60 QKKOTLRLRPELIRANALKTSTFTFLNIMEFINTTCGIGLVSASPMAGMTKMSVQTAAIMVGIIGLFGNGPRLLIAT
 210 220 230 240 250 260 270 280
 65 1170 1200 1230 1260 1290 1320 1350 1380

-2491-

Query: 452 LLAQVKTLLKEKQDILLKGSNMSNLKIVDILE 484
 L+ +VK+L DQILLKGSNMSNL+ +V+ LK
 Sbjct: 421 LVKQVKSLSANDQILLKGSNMSNLMLVESLE 453

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6819> which encodes the amino acid sequence <SEQ ID 6820>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 323/452 (71%), Positives = 387/452 (85%)

- Query: 32 MKLSIHEVAKVVGAKNQSEFQVPLGNIEFDSRNISBGDLFLPLKGARDGHEPIEMAFD 91
 Sbjct: 1 MKL+LHEVAK+V A+N VS+ +DVPL +IEFDSR I+GDFLPLKG RDGHEFI+APF
 MKLTLHEVAKIVDAQNVSDLDVLEHHSFDSRKITGDLFLPLKGQRDGHFIADLAPQ 60
- Query: 92 NGAIATISEKEIBGHFYLLVSDALKAPQVLAQYYIEKQNVDIIVTGSNGRTTKTKMIAA 151
 Sbjct: 61 NG+AT SEKE+ G P+LLV D LKAPQ LA YYI+RM VDIIVTGSNGRT+TKMI A 120
 NGAVATFSEKELPGKPHLLVEDCLKAPQKLAHYIIDMKVDVIATVGSNGRTSTKDMGA 120
- Query: 152 ILSTTYKTYKTQWNYNEIGLPYTVLAMPEDTEKILEMGQDHLGDHVLHSIAKPRIAV 211
 Sbjct: 121 VLSSTYKTYKTQWNYNEIGLPYTVLAMPDTEKILEMGQDHLGDHVLHSIAKPRIAV 180
 +LSTTYKTYKTQWNYNEIGLPYTVLAMP+DTEKI+LEMGQDH+GD1 +LSEIA+PRIAV
- Query: 212 VTLIGEAAHLEFGSREKIAEGKMQITDGMSSDGILLARGDFIIDFPLPANKMTIRFGHDQ 271
 Sbjct: 181 LTLVGEAAHLEFGSRDKIAQGMQITVDGMSDGIILARGDFIIDFPLPANKMTIRFGNQ 240
 +TL+GEAAHLE+FGSR+KIA+GKMQI DGM+SDGILLARGDFIIDFPLP NKM IRFG+ Q
- Query: 272 ELQVTELKEEKHSLLFTKNALEHQLRIPVPGKKNATNANVAAYVGKLLAVABEDIVDALE 331
 Sbjct: 241 EIDVTGIGEDKSLTFTTNVLA TPVSLPLPGKKNATNANVAAYVGKLLAVTDEIIAALQ 300
 E+ VT ++E+K SLTF TN L + +P+PGKKNATNANVAAYVGKLLAV +EDI+ AL+
- Query: 332 NLQLTRNTEWKKANGADILSDVINANPTAMRLILETFSAIPNDGGKKIALLADMKEL 391
 Sbjct: 301 TVTLGNRTWKKANGADILSDVINANPTAMRLILETF+ I N QGKKIA+IADMKEL 360
 + LT NRTEWKK+ANGADILSDVINANPTAMRLILETF+ I N QGKKIA+IADMKEL
- Query: 392 GEQSVDLNQMNSIRPDSIDTILICYQQDIEGLAQLAGNPFPIGRVYFFKQIQVDFQDQ 451
 Sbjct: 361 GKDSVILHSQIDSLTSGNIDQLVFGDHIKEIARLASQVPAEQVHYFLKTKTEQEDFEA 420
 G+ SV LH+Q+I S+ +ID L+ YG I+ LA+LASQ+P +V+P K + +DQF+
- Query: 452 LLAQVKTLLKEKQDILLKGSNMSNLKIVDILE 483
 + V++ L DQILLKGS+SN+L K+VD L
 Sbjct: 421 MAQYVQNILNFPDQILLKGSNMSLEKIVDRL 452

- 50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2209

A DNA sequence (GBSx2328) was identified in *S.agalactiae* <SEQ ID 6821> which encodes the amino acid sequence <SEQ ID 6822>. Analysis of this protein sequence reveals the following:

- 55 Possible site: 17
 >>> Seems to have no N-terminal signal sequence
- 60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1381(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2492-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAC95435 GB:AF068901 D-Ala-D-Ala ligase [Streptococcus pneumoniae]
Identities = 243/346 (70%), Positives = 289/346 (83%)

Query: 3 KRTLLILLYGGRSAERESVLSAESVMRAINVDKFFVKTYFITQVQGFIKTKQEPDEMPSSD 62
K+T+LLYGGRSARERESVLSAESVMRA+NYD+F VKT+FI+Q G FIKTQEF P +
10 Sbjct: 2 KQTIILLYGGRSARERESVLSAESVMRAVNYDRFTVKTFPISQSGIFIKTKQEFSAHQQR 61

Query: 63 EKLMNTQTVLDKRMVRPSDIYDINAIVFPVLHGPMGEDSGIQGLEVLRMYPVGTNILLS 122
++LMTN+T+D K V PS IY++ A+VFPVLHGPMGEDSG+QGFEVL+MPYVG NILLS
10 Sbjct: 62 DRLMTNETIDWKKVAPSAIYERGAIVFPVLHGPMGEDSGVQGLEVLKM+PYVGCNILLS 121

Query: 123 SVAMDKITTKQVLATVGPVQVAYQTYTPEGDDLEHAIKLSLETLSFFPIVFKPANNGSSVG 162
S+AMDKITTK+VL + G+ QV Y ESDD+ I E L++P+F KP+NGGSSVG
15 Sbjct: 122 SLAMDKITTKRVLESAGIAQVVAIVBGGDVDTAKIAVEEKLAYIVFTKPPNNGSSVG 161

Query: 183 SKATDESSLASDAIDALKYDSRLIEQGVYAREIEVGLGNNDVKITFPGEVVKDVPDYD 242
SK+ ++ LR A+ LA +YDSR+L+EGGV AREIEVG+LGN DVK+T PGEVVKDVP DY
20 Sbjct: 182 SKSENQELRQALKAPKYDSRVLEQGVYAREIEVGLGNVDKSTLPGEVVKDVPDYD 241

Query: 243 YDAKYIDNKITMDIPAKVDRAEAMRQYASKAFKAGCLSRCDFFFLTKDQGFINEL 302
YDAKYIDNKITMDIPAK+ + MRQ A AF+AG GLSRCDFF T G+IFINEL
25 Sbjct: 242 YDAKYIDNKITMDIPAKISDDVAVMRQAEATFAIGLGLSRCDFFPYTKGGEIFINEL 301

Query: 303 NTMPGFTQWSMYPLLENMGLTYSDLIEKLVMLAKEMPEKRSHLI 348
NTMPGFTQWSMYPLM+NMG++Y +LIE+LV LAKE P+KRS+HLI
30 Sbjct: 302 NTMPGFTQWSMYPLMDNMGLISYPELIERIVDLAKESFDRKRAHLI 347

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4559> which encodes the amino acid sequence <SEQ ID 4560>. Analysis of this protein sequence reveals the following:

35 Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1451(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 261/348 (75%), Positives = 306/348 (87%)

45 Query: 1 MSKETLLILLYGGRSAERESVLSAESVMRAINVDKFFVKTYFITQVQGFIKTKQEPDEMPSS 60
MSK+TL+LLYGGRSARERESVLSAESVMRA+NYDKF VKTYFITQ+GQFIKQ+F B PS
Sbjct: 1 MSKQTLVLLYGGRSARERESVLSAESVMRAVNYDKLVKTYFITQ+GQFIKTKQSSSEKPS 60

Query: 61 SDEKLMNTQTVLDKRMVRPSDIYDINAIVFPVLHGPMGEDSGIQGLEVLRMYPVGTNILLS 120
E+LMTN+T++L + ++PSDIY+ A+VFPVLHGPMGEDSGIQGLEVLRMYPVGTN+
50 Sbjct: 61 ESEKLMNETIELTKQKPSDIYERGAIVFPVLHGPMGEDSGIQGLEVLRMYPVGTN+ 120

Query: 121 SSSVAMDKITTKQVLATVGPVQVAYQTYTPEGDDLEHAIKLSLETLSFFPIVFKPANNGSSV 180
SSS+AMDKITTK+VL ++G+PQVAY Y+G DLE + +L L+PFI+VFKPANNGSSV
55 Sbjct: 121 SSSVAMDKITTKRVLESIGIPQVAYTYVDGQDLEACLVETLAKLTPPIVFKPANNGSSV 180

Query: 181 GISKATDESSLASDAIDALKYDSRLIEQGVYAREIEVGLGNNDVKITFPGEVVKDVPDYD 240
GISKA + LR AI LAL YDSR+LIEQGV AREIEVG+LGN+ VK+T PGEV+KDVDP
60 Sbjct: 181 GISKAGTKVELRQALKQALTYDSRVLEQGVYAREIEVGLGNDRVKSTLPGEVTKDVDP 240

Query: 241 YDYAKYIDNKITMDIPAKVDRAEAMRQYASKAFKAGCLSRCDFFLTKDQGFIFIN 300
YDY AKY+DNKITM IPA VD++ + MR YA AFKA+G CGLSRCDFFLT+DQG++LN
Sbjct: 241 YDYAKYIDNKITMIPADVDQSIVTMRSYAEVAFKALGGCLSRCDFFLTQDQGVYIN 300

Query: 301 ELNTPGFTQWSMYPLLENMGLTYSDLIEKLVMLAKEMPEKRSHLI 348

-2493-

ELNTMPGFTQWSMYLLWENMGL Y DLIE+LV LA+EMF++RESHLI
 Sbjct: 301 ELNTMPGFTQWSMYLLWENMGLATPOLIELVTLAQEMFDQRESHLI 348

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2210

A DNA sequence (GBSx2329) was identified in *S.agalactiae* <SEQ ID 6823> which encodes the amino acid sequence <SEQ ID 6824>. This protein is predicted to be recombination protein (recR). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.2540(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA044615 GB:U58210 RecN [Streptococcus thermophilus]
 Identities = 181/198 (91%), Positives = 189/198 (95%)

Query: 1 MLYPTPIAKLIDFSKSLPGIGTKTATRLAFTTIGMSDEVDVNEFAKNLLAAKRELYCSVC 60
 Sbjct: 1 MLYPTPIAKLIDFSKSLPGIG KATRLAFTTI MSDEVDV+FAKNLLAAKRELYCSVC 60

Query: 61 GRLTDDDPCLICTDKTRQSVILVVEDSKDVSAMEKIQRYGLYHVLGLISPNKGISPD 120
 G LTTDDPC+ICTD+TRD+ ILVVEDSKDVSAMEKIQRY GLYHVL GLISPNK+ PD
 Sbjct: 61 GRLTDDDPCLICTDTRDKILVVEDSKDVSAMEKIQRYGLYHVLGLISPNKGVGPD 120

Query: 121 DINLKSILTRIMDQVTEVIATNATADGEATSMYISVLKPGAIKVTIRLARGIAGVSDI 180
 DINLKSILTRIMD +V EVI+ATNATADGEATSMYISVLKPGAIKVTIRLARGIAGVSDI
 Sbjct: 121 DINLKSILTRIMDSEVDVIATNATADGEATSMYISVLKPGAIKVTIRLARGIAGVSDI 180

Query: 181 EYADEVILLRAIENRTEL 198
 EYADEVILLRAIENRTEL
 Sbjct: 181 EYADEVILLRAIENRTEL 198

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6825> which encodes the amino acid sequence <SEQ ID 6826>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.2652(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 180/198 (90%), Positives = 192/198 (96%)

Query: 1 MLYPTPIAKLIDFSKSLPGIGTKTATRLAFTTIGMSDEVDVNEFAKNLLAAKRELYCSVC 60
 4LYPTPIAKLIDFSKSLPGIG KATRLAFTTIGMS+KDVN+FAKNLLAAKRELYCSVC+ C
 Sbjct: 1 VLYPTPIAKLIDFSKSLPGIGTKTATRLAFTTIGMSNEVDVFAKNLLAAKRELYCSVC 60

Query: 61 GRLTDDDPCLICTDKTRQSVILVVEDSKDVSAMEKIQRYGLYHVLGLISPNKGISPD 120
 GRLTDDDPCLICTD+RDQ+ ILVVED+KDVSAAMEKIQRY+G YHVLGLISPNK+ PD
 Sbjct: 61 GRLTDDDPCLICTDTRDQTTILVVEDAKDVSAMEKIQRYGYHVLGLISPNKGVGPD 120

-2494-

Query: 121 DINKSLITRLMDGQVTEVIVATNATADGEATSMYISRVLPKPGIKVTRLARGIAGVSDI 180
 DINKSLITRLMDG+V+EVIVATNATADGEATSMYISRVLPKPGIKVTRLARGIAGVSDI
 Sbjct: 121 DINKSLITRLMDGKVEVIVATNATADGEATSMYISRVLPKPGIKVTRLARGIAGVSDI 180

5 Query: 181 EYADEVTLLRAIENRTEL 198
 EYADEVTLLRAIENRTEL
 Sbjct: 181 EYADEVTLLRAIENRTEL 198

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2211

A DNA sequence (GBSx2330) was identified in *S. agalactiae* <SEQ ID 6827> which encodes the amino acid sequence <SEQ ID 6828>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3144 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2212

A DNA sequence (GBSx2331) was identified in *S. agalactiae* <SEQ ID 6829> which encodes the amino acid sequence <SEQ ID 6830>. This protein is predicted to be penicillin-binding protein 2b. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL likelihood = -13.69 Transmembrane 23 - 39 (17 - 46)

----- Final Results -----
 bacterial membrane --- Certainty=0.6477 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAC44614 GB: U58210 penicillin-binding protein 2b [Streptococcus thermophilus]
 Identities = 341/683 (49%), Positives = 477/683 (68%), Gaps = 12/683 (1%)

Query: 4 RKKRYRLTVKKQASIPRELMLLPFIIVLAFVLIIRLQCMQIGQGSFYMKKLIALTSYTT 63
 ++K R ++ +I RR+ ILF ++ +LF +L RL MQ+ +SPY KKL + YT
 Sbjct: 18 KRKEKRAKPKPKFVNISRRVYLLPGVFVFLFLLLPARLTYMQVYKNSFYTKKLDNSKYT 77

Query: 64 VKESKARQIPDAKGVVLENDERPTVAFSRGNISQSIKHLANKLSHYITLTFVASSD 123
 V+ + RQIPDAK+ L N + + F+R N +SS ++K +A +L+ +TIFR +D
 Sbjct: 78 VRIASERQIPDAKGIALTQKSDVITFTSNLVSDDIMKSVARLATINLTETKVID 137

Query: 124 RAKRDYLLADKANYKKVVEISLPDSKRYDKFGNHLARSTVYANAVAAVFPVAINSEDELK 183
 R KR++YLAD ANYK+VV LP+ K+ DRFGN LAR+T+Y NA+ AVP A++YSEDELK
 Sbjct: 138 RQKREFYLADSANYYKRVNDLNDKKITKPGNKLARATTYNNAINAVPDAVDYSEDELK 197

-2495-

Query: 184 VVALFVQNNATPTFGSVKLSSTGRLSDQKIKLADAKGKRLGISVTSNWHRRKGTSLSDI 243
+V +++ MSA F +V L T +L+ DCI + A +KEL GI V +W R +SLS +
Sbjct: 198 IVYIYSHNAGVSNFSTVILKTDLTDPQIAVAAKQKELNGIRVAKDWERHTSDSSLSPL 257

5 Query: 244 LGTISTEKGALPREVEKYLKGYSLNDRVGTSTYLEKQYVEDLOGIRQIRVVVWKKGV 303
+G +S+ +AGLP+E+ K YLKKGY+LNDRVGTSTYLEK+YE++LQG +R++ V+K+GKV
Sbjct: 258 IGRVSSSSAGLPQEDAKDYLLKGYSLNDRVGTSTYLEKYEELQGHVREITVYDSEKV 317

10 Query: 304 VSDNITQEGKSGRNKLITDLAYQNKVSTILKQYVSSKSSGRASPSEGMVYAVIEPSTG 363
SD I Q+G G NLKLTIDL+Q VE IL Q SE+S +A+SEGMVAV +TG
Sbjct: 318 DSDKIQKSGKRNKLITDLDFQKGVEDILQQLSSSEISGNKATYSEGMVAVVMDATG 377

15 Query: 364 KVLAMAGLNNDHG--NLVDLSLTIAKNFTFGSVKGTSLSSGWNKVLKQNEVLYDQEI 421
VLAMG K++ G + D+LGTI FTGSGVVKGTIL++G+ + G++VL DQ I
Sbjct: 378 AVLAMAGQKHEGGAQDFKADNLGTTITDFTFGSVVKGTILTAGNRSAGATYGDQVLIDQPI 437

20 Query: 422 -----ANIRSWFT-RGLTPISAAQALEYSSNTVMQVVALRMQDDINTGDLTDRGYQEA 475
I SWFT +G I+A QALEYSSNTVMQV++ ++Q Y G I++ ++A
Sbjct: 438 NIASSPPTISWFTDKGSRATITATQALEYSSNTVMQVIAIKRLGQQVFGMSLSTDMEKA 497

25 Query: 476 MAKLRKTYGEGYGLGVSTGLDLP-ESBGYPVGKYSGLTIMSEFQGDAYTPMOLQGYIST 534
M LR TY E+G+GVSTGLDLP ESBGY+P Y++ L E+FGQD+YT +Q QY++
Sbjct: 498 MTTLEDYAEFGMGVSTGLDLPGBESGYIPGVNANVLTAPGQYDYTTIQLAQYVAS 557

30 Query: 535 IANGNRLAPHVSDIYSDNENKPAOLVSTIPKTIANKIAISDQELAIQGFYFVNS 594
IAN G R+APH+V IY+ + L ++ +LAK+++ ++L IIQ+GF++VNS
Sbjct: 558 IANGGKRVAPHIVOGIYDAGKNGSLTSLSTVDTRVLNKLSDSKQLAITQQGFMDVNS 617

35 Query: 595 GSGYATGTSMRGVTTISGKTGTATETPAKNVQGVSTYVNLMAIYDINR---KIAVAVM 651
GS ATG +M ++ ISGKTGTATET+A +G +V+T NLNA+AY T + K+AV +M
Sbjct: 618 GSSLATGKMASSIIPIGKTGTATETVATDGSNSVITVNLNAVAYATADQTKLAVGIM 677

Query: 652 YPHVTTDTTKSHQLVARIMDQY 674
YPH +K+HQ + +++ Y
Sbjct: 678 YPHALDNKSKAHQNAVKAIMELY 700

A related GBS gene <SEQ ID 8997> and protein <SEQ ID 8998> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop Possible site: -1 Crend: 8
McG: Discrim Score: -12.38
GVH: Signal Score (-7.5): -5.9
Possible site: 35
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -12.42 threshold: 0.0

45 INTEGRAL Likelihood =-12.42 Transmembrane 23 - 39 (18 - 46)
PERIPHERAL Likelihood = 4.56 355
modified ALOM score: 2.98

50 *** Reasoning Step: 3
----- Final Results -----
bacterial membrane --- Certainty=0.5967 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50.5/71.3% over 683aa
thermophilus
60 GP|1685112| penicillin-binding protein 2b Insert characterized
ORF02276 (307 - 2322 of 2643)
GP|1685112|gb|AAC44614.1||U59210(17 - 700 of 704) penicillin-binding protein 2b
{Streptococcus thermophilus}
65 %Match = 38.5
Streptococcus

-2496-

%Identity = 50.4 %Similarity = 71.2

Matches = 342 Mismatches = 189 Conservative Sub.s = 141

```

5      108      138      168      198      228      258      288      318
MNGR*NS*LPTTCFRI**KIKPCFRILLR*II*SLYKKFRPSWLEFFIITNLSVCKKFL*YNSQSPYSKELMLNRKK
                                         :   :   :   :   :
MTSFWEKNSQKWKWRQKKR
                                         10      20

10     348      378      408      438      468      498      528      558
RYRLTVKQKQASIPRRINLLFFIIXLLFTVLILRLBQMOIGQCSFYMKKIATLTSYTVKSKARQQIFDANGVVLVND
|   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
EKRAKPKPKFVNISSRYLLFGVVVFLFLLPARLTYMQVYNKSPYTKKLEDSKYTVRIASERQIQFANGIALTYNQS
                                         30      40      50      60      70      80      90      100

15     588      618      648      678      708      738      768      798
RPTVAFSRGNTISQSIKELANKLSHYITLTVASSDRAKRDYTLADIKVYKVVESLPDSKRYDKPGNHLAESTVYANA
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
KDVITPRLSHLVSDTMKSVARLATALVLTETKVTDRQKRBFHLADSANYKRVNDLPNDKIKTKPGNKLAEATTYNA
                                         110     120     130     140     150     160     170     180

20     828      858      888      918      948      978      1008     1038
VAAVPSAINTSEDELKVALFVQWQAT*TPGSKVLSTGSLDDQIKKLADKKELGLSVTSHWRKRLSTLESDILQ
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
INAVPDEAVSEDELKIVITYSHRRAVSNFSTVILKTADUTPDQIAIYAAQKSLANGIRAKDWEHSTSDSLPLIGR
                                         190     200     210     220     230     240     250     260

25     1068     1098     1128     1158     1188     1218     1248     1278
ISTEKAGLPREEVKKYKKGYSLNDVRGTSYLEKQYEDDLQGIQRKVVVYKKGKVVSDNITQBGSKSQRKLGLTIDNY
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
VSSSEAGLPQEDAKDYLLKGYALNDVRGTSYLEKEYEELQSKHTRVREITVQEKGVDSDKIIQKSKSNMLKLTIDLDF
                                         270     280     290     300     310     320     330     340

30     1308     1338     1368     1398     1428     1452     1482     1512
QNKVESILKQYVSELSGRASPSGMYAVAIKPSGTGKVLAMAGLNDHG--NLVDSLSGTIAKPIPGSVVKGATLSSG
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
QKGVESILGQQLSSISGNKATYSBGMVAVVMDTGAVALMAGQKHQGAQDFKADALGTTITDPTPGSVVKGATLFG
                                         350     360     370     380     390     400     410     420

35     1542     1566     1587     1614     1644     1674     1704     1734
WENKVLGRNEVLYDQ--EIAN--IRSWFT-RGLTPISSAAQALYSSNTVMQVALLMGQOYVNGGALTIRGYQEMAK
:   :   :   :   |   :   :   :   |   :   :   :   |   :   :   :   |   :   :   :   |   :   :   :
WRSGLAYGDOVLTDQPINIASSPPTISWFTDKGSRATITATQALYSSNTVMQVLAIRLGGQYVPG*SLSTDNMEKAMIT
                                         430     440     450     460     470     480     490     500

40     1764     1821     1851     1881     1911     1941     1971
LRKTYGEGYGLVSTGLDLP-RSGYVPGKYSLOTTLMESRGQYDAYTPMQLQVISTANNKRLAPHVSDIYBGRDSE
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
LRDTYAEFGVSGVSTGLDLPESBGYIIFKNYVANVLTFAFGQYDSYTTIQLAQVSIANGGKRVAPHVGGIYDAGKNG
                                         510     520     530     540     550     560     570     580

45     2001     2031     2061     2091     2121     2151     2181     2211
KFAQLVRSITPXTKNTAKIASDQELAIQEGFYNNVNGSGGYATGTSMRGNVITTSKGTGTAETFAKIVNGQTVSTYNLNA
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
SLGTLSTVDTVKLNKLSLDSKQLGIQOQFHDVNNSSSLATGKAMASSIIPISGKTGTATGYATDQSGNSVITYNLNA
                                         590     600     610     620     630     640     650     660

50     2262     2292     2322     2352     2382     2412     2442
IAYDTNR---KIAVAVMYPEVTTDTTKSHOLVARDMIDQYISQFTGQ*ERTPECFCTQGLLN*LTAPQNYV*VLKQGVH
|   |   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
VAYATAKDGTKLAVGIMYPHALDWKSKAHQNAVKAIMELYNTH
                                         670     680     690     700

```

SEQ ID 8998 (GBS292) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 9; MW 103kDa).

GBS292-GST was purified as shown in Figure 211, lane 7.

-2497-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2213

A DNA sequence (GBSx2332) was identified in *S.agalactiae* <SEQ ID 6831> which encodes the amino acid sequence <SEQ ID 6832>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2644 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>P:CB51328 GB:A7131985 phosphoglyceromutase [Streptococcus pneumoniae]
Identities = 219/230 (95%), Positives = 226/230 (98%)

Query: 1 MVKLVFARHGSESWNKANLFTGWADVLSERGTQQAIDAGKLIQAGIEFDLAFTSVLKR 60
MVKLVFARHGSESWNKANLFTGWADVLSERGTQQAIDAGKLI+ AGI+FD A+ISVLKR
Sbjct: 1 MVKLVFARHGSESWNKANLFTGWADVLSERGTQQAIDAGKLIKAGIKFDQAYTSVLKR 60

Query: 61 AIKTTNLALEAADQLWVFEKSWRLNERHYGSLTGKNKASABQFQDEQVHINRSIDVL 120
AIKTTNLALEA+DQLWVFEKSWRLNERHYGSLTGKNKASABQFQDEQVHINRSIDVL
Sbjct: 61 AIKTTNLALEAADQLWVFEKSWRLNERHYGSLTGKNKASABQFQDEQVHINRSIDVL 120

Query: 121 PPMMAKDDHSAHTDRRYASLDDSVIPDAENLKVTILERALPFMEDKIALPALDKGNVFG 180
PP+M +DDEHSANTDRRYASLDDSVIPDAENLKVTILERALPFMEDKIALPALDKGNVFG
Sbjct: 121 PPMMAKDDHSAHTDRRYASLDDSVIPDAENLKVTILERALPFMEDKIALPALDKGNVFG 180

Query: 181 AHGNSIRALVGHIKQLSDDSEINDVEIPNFPPLVFEFDEKINLVSEYILGK 230
AHGNSIRALVGHIK L SDDSEINDVEIPNFPPLVFEFDEKINLVSEYILGK
Sbjct: 181 AHGNSIRALVGHIKQLSDDSEINDVEIPNFPPLVFEFDEKINLVSEYILGK 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6833> which encodes the amino acid sequence <SEQ ID 6834>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2646 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 206/229 (89%), Positives = 214/229 (92%)

Query: 1 MVKLVFARHGSESWNKANLFTGWADVLSERGTQQAIDAGKLIQAGIEFDLAFTSVLKR 60
MVKLVFARHGSESWNKANLFTGWADVLSERGTQQAIDAGKLI+ AGIEFDLAFTSVL R
Sbjct: 1 MVKLVFARHGSESWNKANLFTGWADVLSERGTQQAIDAGKLIKAGIKFDLAFTSVLKR 60

Query: 61 AIKTTNLALEAADQLWVFEKSWRLNERHYGSLTGKNKASABQFQDEQVHINRSIDVL 120
AIKTTNLALE A QLWVFEKSWRLNERHYG LTGKNKASABQFQDEQVHINRSIDVL
Sbjct: 61 AIKTTNLALENAQQLWVFEKSWRLNERHYGALTGKNKASABQFQDEQVHINRSIDVL 120

Query: 121 PPMMAKDDHSAHTDRRYASLDDSVIPDAENLKVTILERALPFMEDKIALPALDKGNVFG 180
PP MAKDE+SAH DRRYA ID ++IPDAENLKVTILERALPFMEDKIALPALDKGNVFG
Sbjct: 121 PPMMAKDDHSAHTDRRYADLDPAIPDAENLKVTILERAMPVWEKIALPALDKGNVFG 180

Query: 181 AHGNSIRALVGHIKQLSDDSEINDVEIPNFPPLVFEFDEKINLVSEYILGK 229

-2498-

AHNSIRALVKHIK LSDSDIMDVETPNFPLVPE DKLN+V EYVLG
 Sbjct: 181 AHNSIRALVKHIKGLSDSDIMDVETPNFPLVPELDKLNIVKEYLGLG 229

SEQ ID 6832 (GBS110) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 8; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 10; MW 53.9kDa).

The GBS110-GST fusion product was purified (Figure 204, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 252A), FACS (Figure 252B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2214

A DNA sequence (GBSx2333) was identified in *S.agalactiae* <SEQ ID 6835> which encodes the amino acid sequence <SEQ ID 6836>. This protein is predicted to be triosephosphate isomerase (tpiA). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 36 - 52 (36 - 52)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the sequences in the GENPEPT database.

>GP:AAC43268 GB:U07640 triosephosphate isomerase [Lactococcus
 lactis]
 Identities = 164/252 (65%), Positives = 202/252 (80%)
 Query: 1 MSRKPFIAQNWKNKPNPEEAKPIEAVASKLPSSSELVEGIAAPALITLSTVLEAAKGSGL 60
 MSRKPF IAGNWKNK ER+AF+EAV + LPSS+ VE+ I APAL L+ + +GSEL
 Sbjct: 1 MSRKPIIAGNWKNKTLSEQAQFVEAVKNNLPSSNVSIGAPALFLAPVAYLRQSGSL 60
 Query: 61 KIAAQNSYFENGAGFTGENSPKVLAMGTQYVVIHSERRDYFHETDQDINKKAKAIFAN 120
 K+AA+NSYFEN+GAFTGENSP + ++G +Y++IGHSERR+YFHETD+DINKKAKAIFA
 Sbjct: 61 KIAAENSYFENAGAGFTGENSPAATVDLQIEYIIIGHSERRDYFHETDQDINKKAKAIFA 120
 Query: 121 GLTPIICCGRSLETYEAGKAVFVGQVSAALAGLSEBQVSSLVIAIEPIWAIGTKSAT 180
 G TPI+CCGE+LET+EAGK E+V Q+ A LAGL+ EQVS+LVIAIEPIWAIGTK+AT
 Sbjct: 121 GATPIICCGETLETYEAGKATWVWSQTEAGLAGITABQVSNLVIAIEPIWAIGTKTAT 180
 Query: 181 QDDAQNMCKAVRDVVAADFQGAQVADKVRVOYGGSVKPNVAEYMACPDVDGALVGGASLE 240
 + A C VR V +G+ V++ VR+QYGGSVKPR + MA ++DQALVGGASLE
 Sbjct: 181 NEIADETCGVVRSTVEKLYGKEVSEAVRIQYGGSVKPRTEGLMAKENIDGALVGGASLE 240
 Query: 241 AESFLALLDFVK 252
 A+SFLALL+ K
 Sbjct: 241 ADSFLALLEMYK 252

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6837> which encodes the amino acid sequence <SEQ ID 6838>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

-2499-

INTEGRAL Likelihood = -1.81 Transmembrane 36 - 52 (36 - 52)

----- Final Results -----

5 bacterial membranes --- Certainty=0.1723 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 220/251 (87%), Positives = 237/251 (93%)

10 Query: 1 MSRKPTIAGNWKMKNPQERAKAFIEAVASKLPSELVEAGIAAPALITLSTVLEAAKGSLE 60
MSRKP IAGNWKMKNP+ERAKAF+EAASKLP++LV+ +AAPA+ L T +EAAK S L
Sbjct: 1 MSRKPTIAGNWKMKNPQERAKAFIEAVASKLPSTLDVVAAPAVDLVTTEAAKDSVL 60

15 Query: 61 KIAAQNSYFENGGAFTGENSPKVLAEKMDYVVIHGSRRDYFHETDQDINKKAKAIFAN 120
K-AAQN YFEN+GAFTGS SPKVLAEKMG DYVVIHGSRRDYFHETD+DINKKAKAIFAN
Sbjct: 61 KVAQNICYFENTGAFTGTSPKVLAEKMGADYVVIHGSRRDYFHETDSDINKKAKAIFAN 120

20 Query: 121 GLTPIICCGSELETYEAGKAVEFVGAQVSAALAGLSBEQVSSLVIAYEPIWAIGTKSAT 180
GLTPI+CCGSELETYEAGKAVEFVGAQVSAALAGLS EQV+SLV+AYEPIWAIGTKSAT
Sbjct: 121 GLTPIVCCGSELETYEAGKAVEFVGAQVSAALAGLSAQVSLVLAYPEIWAIGTKSAT 180

25 Query: 181 QDDAQNMCKAVRDVVAADFQGVADKVRVQYGGSVKPEINVAACPDVDGALVGGASLE 240
QDDAQNMCKAVRDVVAADFQGVADKVRVQYGGSVKPEINVAACPDVDGALVGGASLE
Sbjct: 181 QDDAQNMCKAVRDVVAADFQGVADKVRVQYGGSVKPEINVAACPDVDGALVGGASLE 240

Query: 241 AESFLALLDFV 251
A+SFALLDF+
Sbjct: 241 ADSFLALLDFL 251

30

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2215

35 A DNA sequence (GBSx2334) was identified in *S.galactiae* <SEQ ID 6839> which encodes the amino acid sequence <SEQ ID 6840>. Analysis of this protein sequence reveals the following:

Possible site: 23
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3050 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AA541198 GB:U75481 elongation factor-Tu [Streptococcus mutans]
Identities = 44/45 (97%), Positives = 45/45 (99%)

50 Query: 1 MVMFGDNVTIEVELIHPIAVEQGTTFSIREGGRTVGGGIVSEIEA 45
MVMFGDNVTI+VELIHPIAVEQGTTFSIREGGRTVGGGIVSEIEA
Sbjct: 117 MVMFGDNVTINVELIHPIAVEQGTTFSIREGGRTVGGGIVSEIEA 161

There is also homology to SEQ ID 1022:

Identities = 44/45 (97%), Positives = 44/45 (97%)

55 Query: 1 MVMFGDNVTIEVELIHPIAVEQGTTFSIREGGRTVGGGIVSEIEA 45
MVMFGDNVTI+VELIHPIAVEQGTTFSIREGGRTVGGGIVSEIEA
Sbjct: 371 MVMFGDNVTINVELIHPIAVEQGTTFSIREGGRTVGGGIVSEIEA 415

-2500-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2216

A DNA sequence (GBSx2335) was identified in *S.agalactiae* <SEQ ID 6841> which encodes the amino acid sequence <SEQ ID 6842>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -2.66   Transmembrane  81 - 97 ( 80 - 97)
      INTEGRAL Likelihood = -2.60   Transmembrane  18 - 34 ( 17 - 34)

----- Final Results -----
      bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2217

A DNA sequence (GBSx2336) was identified in *S.agalactiae* <SEQ ID 6843> which encodes the amino acid sequence <SEQ ID 6844>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0596(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2218

A DNA sequence (GBSx2337) was identified in *S.agalactiae* <SEQ ID 6845> which encodes the amino acid sequence <SEQ ID 6846>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3559(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2501-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2219

A DNA sequence (GBSx2338) was identified in *S. agalactiae* <SEQ ID 6847> which encodes the amino acid sequence <SEQ ID 6848>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF96266 GB:AB004374 hypothetical protein [Vibrio cholerae]
Identities = 56/167 (33%), Positives = 89/167 (52%), Gaps = 12/167 (7%)

Query: 18 LAIIKSLPLNDOWLCAGTLANFVWNLKLS-GINETILTSDDIVVFFDKNI---SYESTVWLE 73
L + L L C+ AG +RN VM+ L + T +DIDV+FFD + YE++ LE
Sbjct: 41 LECVYQLELPCQCYAAGFVFNLVWDSLHINVKLTPLND:DVIFFDADCLDSITYEKS--LE 98

Query: 74 QQLKONYPCQYDWELAKNEFYMNTHSPNTPKTYSSKDAISKFFPKCTAWGARLDDRNQLELY 133
+L + EQ +W+++KH+ M+ + + F Y S+ DR+S +PEK TAV R + ++ E
Sbjct: 99 LKLSEQMPQLNMQVKNQAKHQLQNGDNP-QYSTLDAMSTWPEKETAVAVRKEVHDREYECI 157

Query: 134 LPYGEELINPVSPTFYFEEDLLRYNVTLXRVKDKHGWNNINPLTI 180
+G E + ++ P Y ++ RV K W +WF L I
Sbjct: 158 SAFPGEELPFGGFITHNP-----KRAYGIFENRVKSGWLAAMVFNLR 199
```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2220

A DNA sequence (GBSx2339) was identified in *S. agalactiae* <SEQ ID 6849> which encodes the amino acid sequence <SEQ ID 6850>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2779 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13060 GB:Z99110 yjdf [Bacillus subtilis]
Identities = 47/138 (34%), Positives = 93/138 (67%), Gaps = 2/138 (1%)

Query: 1 MKNTVYYPGNFWLGLI EYDDGDIYKVFYFPGKEPKDDVFNFINHKLMDLKKYEFVKI 60
MK+T+Y+DG FW+G++E D+G + FR+ PGKEP+D +V F++++L +++ + E +
Sbjct: 24 MKLTITYYDQGFVWGVVGVVFNGLKRAPRHLPGKEPRDSEVLFVFNQLNMAQAE--QB 81

Query: 61 DISLKRTNHHKKS PKRMOREINREKKKPVVSTKAQLAMKTHMS IKNRQLSQKCKKNEL 120
+ Lt + K +PKR+QR+++E + V++KQ A+K + K +++ K ++ +
Sbjct: 82 GVRLOGRROKKNPKRLCROVSKELKNAGVTSKAQEAIKLELEARKQKKQKQIMKEGREV 141
```

-2502-

Query: 121 RGHRYQLNQEKRYQKKKG 138
 ++ RY LK++K +K +G
 Sbjct: 142 KEGYMLKKQAKKKHRG 159

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2221

- 10 A DNA sequence (GBSx2340) was identified in *S.galactiae* <SEQ ID 6851> which encodes the amino acid sequence <SEQ ID 6852>. This protein is predicted to be ComX1. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3143 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 9469> which encodes amino acid sequence <SEQ ID 9470> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD50429 GB:AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 61/152 (40%), Positives = 95/152 (62%)

- 25 Query: 5 BELFDKVPKIPVIMKLARNYFVQLMEYDWDIQGRVLFRLLEEHPYLLDNESKLPYIFKTK 64
 +EL+++V+ V K R Y+ LWE DW QEG + L L+ L+D+ +L YFKTK
 Sbjct: 3 KELYEEVQGTIVYKCRNYTLLHMLSDWDQEGMLCLHELISREBGLVDDIPRLKPYFKTK 62
- 30 Query: 65 FSNYLADVLRHQDQKRPQKHPYEEISGVSHYVSKGLVLDYIAYEDTLKVEETLSD 124
 F N + D +R Q+ QKR+++K FYEE+ E+SH + GL LDDY + +TL S
 Sbjct: 63 FNRILDYIRKQESQKRRYDKEPYEEVGEISHRISBGGMLLDYVLFHETLRDYNKQSK 122
- 35 Query: 125 IDKEKPEHLISGERFAGKKQFIRDQFFPNAF 156
 +E+ E++S ERF G+++ +RD++ F F
 Sbjct: 123 EQKEELERLVSNRFRGRQVRLDLRIVKPEF 154

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6853> which encodes the amino acid sequence <SEQ ID 6854>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-10.35 Transmembrane 9 - 25 (7 - 28)

- 45 ----- Final Results -----
 bacterial membrane --- Certainty=0.5140 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 50 A related sequence was also identified in GAS <SEQ ID 9163> which encodes the amino acid sequence <SEQ ID 9164>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-10.35 Transmembrane 2 - 18 (1 - 18)

-2503-

----- Final Results -----

5 bacterial membrane --- Certainty=0.160 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: AAD50429 GB: AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 60/149 (40%), Positives = 98/149 (65%)

10 Query: 41 FEKVKPIILKLRHYIQLWDRDNLQBGHIIQLLERYPELIEREERINRYPTKFPSS 100
 +E+V+ + K + YY+ LW+ DN QEG + L +L+ R L+++ RL +YPTKFP +
 Sbjct: 6 YBEVGTGTYKCRNEYTLHLWELSDNQBGLCLHGLISKRKGLVDDIPRLRYKFKTKPRN 65

15 Query: 101 YLNDLLRRQESQKQKPHKLAYEISGEVAHAIPSRGLWDDYVAYQEVIASLENQINSQER 160
 + D +R+QESQK++ K YEE+GE++H I GLWDDY + E + N+ + +++
 Sbjct: 66 RILDYIRKQSKKRYDKPEYEEVGEISHRISBGLWDDYLPHEITRDYNNQSKSKQ 125

20 Query: 161 MQFQALIRGERFGRKALLRKISPYKFKEF 189
 + + ++ ERF+GR+ +LR + FKEF
 Sbjct: 126 EBLERVLNERNFRGRQVLRDLRIVFKEF 154

An alignment of the GAS and GBS proteins is shown below.

Identities = 78/149 (52%), Positives = 116/149 (77%)

25 Query: 8 FDKVKPIVMKLRNRYFVLWEYDDWICQGRIVLFRLEBHPYLLDNESKLPFYPTKFPN 67
 F+KVKPI++KL+R+Y++QLW+ DDW+QEG I+L +LE +P L++ E +L+ YPTKFP+
 Sbjct: 41 FEKVKPIILKLRHYIQLWDRDNLQBGHIIQLLERYPELIEREERINRYPTKFPSS 100

30 Query: 68 YLNDVLRHQCCQKQKPHKLAYEISGEVAHAIPSRGLWDDYVAYQEVIASLENQINSQER 127
 YL DL+R Q C OKRQ+K+ YERI EV+H + S+GL LDDY+AY++ + +E L+ ++
 Sbjct: 101 YLNDLLRRQESQKQKPHKLAYEISGEVAHAIPSRGLWDDYVAYQEVIASLENQINSQER 160

35 Query: 128 BKPEKLISGERFAGKQCFIRDIQPFNFAP 156
 +P+ LI GERF G++ +R I P+F F
 Sbjct: 161 MQFQALIRGERFGRKALLRKISPYKFKEF 189

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 **Example 2222**

A DNA sequence (GBSx2341) was identified in *S. agalactiae* <SEQ ID 6855> which encodes the amino acid sequence <SEQ ID 6856>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 45 INTEGRAL Likelihood = -2.23 Transmembrane 166 - 182 (166 - 182)

----- Final Results -----
 bacterial membrane --- Certainty=0.1893 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database.

>GP: CA99510 GB: Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 57/226 (25%), Positives = 97/226 (42%), Gaps = 22/226 (9%)

55 Query: 4 VRLYIARHGKTMPTNIGRAQQGWSDTPLTFGLGELGKLGGLGKASNISFKEAFSSDSGR 63
 +RL+I RHG+T N QG DT + GE +LQ L++ I F+ SSD R
 Sbjct: 17 IRLFIIRHQTEHNKVKILQGHKDTNINPTGEQATKLGHYLRSGIHFDKVVSSDLKRC 76

-2504-

Query: 64 LQTMETILREVVQQRNIPYTRDKRIREWCPSGLGGYDGLFNGVLPRVSNMGDSHMTHEE 123
 QT ++L+ +QEN+P + +RE G ++G M E+
 Sbjct: 77 RQTALVLKHSKQNVPTSYTSGLRERYMGVIEG-----MQITEAEK 118

5 Query: 124 IANLIQVDTAGWAEPAWILSNRILSGPTAIKAKKIEDIGGNAIVVSHGMTIATFL-WL- 181
 A+ + + E R+ + + G N +VSHG I L WL
 Sbjct: 119 YADKHGEGSFRNFGKSDDFVARITGCVSEEVAEASNEGVKNILALVSHGGAIRMLQLWL 178

10 Query: 182 IDEHTPRSLGLDNGSVSVDF--EDGTFSIQSIGDMSYREKGRILE 225
 ++ + + N SV+VD+ + F ++ +G+ + G ++
 Sbjct: 179 YRNHQAHIKIVPTSVTIVDYVKCKQFIVRRVGNQHLGDGEFVV 224

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6857> which encodes the amino acid sequence <SEQ ID 6858>. Analysis of this protein sequence reveals the following:

15 Possible site: 57

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.69 Transmembrane 170 - 186 (170 - 186)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:CAA99510 GB:Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 64/231 (27%), Positives = 98/231 (41%), Gaps = 27/231 (11%)

30 Query: 5 RLYIARHGKTMPTNIGRAQGSWDTPLTKKGEGIRELGLGLKDAITPFKAAPSSDSGRIM 64
 RL+I RHG+T N OG DT + GEE +LG L+ I F SSD R
 Sbjct: 18 RLFIIRHGQTEHNKKILQSHKDTNINIVGEEQATKLGHYLRSGIHFKVVSDDLKRCR 77

Query: 65 QTIEILIRESENEFLPYTKDNRIREWCPGSLGGTYDSELFGLVLPRTKAFENRDNLDVDP 124
 QT ++L+ S+ E +P + + +RE G +EG +E
 35 Sbjct: 78 QTALVLKHSKQNVPTSYTSGLRERYMGVIEG-----MQITEA----- 116

Query: 125 YSELAESIVEVDANNAEPMEVLKRIWGFPAIALSIQAGGNGALVSHGMTITGTITFL- 183
 + A+ E N+ E + R+ E N G N +VSHG I L
 40 Sbjct: 117 -EKYADKHGEGSFRNFGKSDDFVARITGCVSEEVAEASNEGVKNILALVSHGGAIRMLIQ 175

Query: 184 WL--IDPDRDKQYIDNGSVTVVEE--DDGQFTIKTGDMSYRYGRRIIEE 230
 WL + K + N SVT+V+ D QF ++ +G+ + G ++ +
 Sbjct: 176 WLKYENHQAHIKIVPTSVTIVDYVKCKQFIVRRVGNQHLGDGEFVVD 226

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 150/231 (64%), Positives = 182/231 (77%), Gaps = 5/231 (2%)

Query: 1 MSKVRLYIARHGKTMPTNIGRAQGSWDTPLTFQELGIKELGLGLKASNISFKEAPSSDS 60
 M+K RLYIARHGKTMPTNIGRAQGSWDTPLT GE GI+ELGIGLK + I FK AFSSDS
 50 Sbjct: 1 MTKRLYIARHGKTMPTNIGRAQGSWDTPLTKKGEGIRELGLGLKDAITPFKAAPSSDS 60

Query: 61 GRTLQTMETILREVVQQRNIPYTRDKRIREWCPSGLGGYDGLFNGVLPRV-----SNGDM 116
 GRT+QT+ETILRE + E +PT+D RIREWCPSGL+G YD +LF GVLER + ++
 55 Sbjct: 61 GRTMQTIEILIRESENEFLPYTKDNRIREWCPGSLGGTYDSELFGLVLPRTKAFENRDNLD 120

Query: 117 SHLTHEEIANLIQVDTAGWAEPAWILSNRILSGPTAIKAKKIEDIGGNAIVVSHGMTIA 176
 + + E+A I +VDTA WAEFW +L RI GF AIA I++ GGGNA+VSHGMTI
 Sbjct: 121 RDVPYSELAESIVEVDANNAEPMEVLKRIWGFPAIALSIQAGGNGALVSHGMTIG 180

60 Query: 177 TFWLWIDHSTPRSLGLDNGSVSVDFEDGTFSIQSIGDMSYREKGRILEK 227
 TFWLWID + +DNGSV+VV+F+DG F+I+I+IGDMSTR +GREI+E+
 Sbjct: 181 TFWLWIDFDRKQY-IDNGSVTVVEFDDGQFTIKTGDMSYRYGRRIIEE 230

-2505-

A related GBS gene <SEQ ID 8999> and protein <SEQ ID 9000> were also identified. Analysis of this protein sequence reveals the following:

Cytoplasmic predicted but experimentally found on the surface of *Streptococcus*

```

5      32.3/52.0% over 184aa
                                     Thermotoga maritima
      EGAD|165681| phosphoglycerate mutase Insert characterized
      GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase Insert characterized
10     FIR|G72260|G72260 phosphoglycerate mutase - (strain MSB8) Insert characterized

      ORF01265(268 - 870 of 1248)
      EGAD|165681|TM1374(1 - 185 of 201) phosphoglycerate mutase (Thermotoga maritima)
      GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase (Thermotoga maritima)
15     FIR|G72260|G72260 phosphoglycerate mutase - Thermotoga maritima (strain MSB8)
      %Match = 6.3
      %Identity = 32.2 %Similarity = 52.0
      Matches = 57 Mismatches = 78 Conservative Sub.s = 35

20     105      135      165      195      225      255      285      315
      RGRNNSYELIFNPFMSLLKRLNRYFPCSR*LQNFFIGIKV*YIPVKAFVFCYNILKCL*GVSMKSVRLYIARHGKTMFNII
                                     ::||: |||::|
                                     MKGLYLIRHGETIWNKK
                                     10

25     345      375      405      435      465      495      519      549
      GRAQGWSDTPLTFGELGIKELGLKASNISFKKAPSSDSXRTLQTMELILREVQQENI--PYTRDKRIKRE*CFGSLDG
      | ||: || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      GLWQGV*DVPLAEREGEQARKLANSLK----RVDALYSPLKRSLETAERIPKBIIVZEDLRRCRISLA-----
30     30      40      50      60      70      80

      579      609      639      669      699      729      759
      VYGDGLFNGVLPRVNSGDSMLTTHEETIANLICQVDTAGWA-----EPWAILSNRILSGFTAIKKEDIGGGNAI
      : || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      -----NGLTVEE-AIREYPVEFKWSSDPNFGHGLESMEVQNRVVKAIMKIVSQEKLNGSENVV
35     90      100      110      120      130      140

      789      816      840      870      900      930      960      990
      VVSHGNTIATFL-NLIDHST--PRSLGLDNGSVVDVFDQTFPSIQSGIDMSYREKGREILEKTLQ*KKIKLSDSV*LVF
      : || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
40     IVSHSLSLRAFCWILGLPLYLHNRNFKLDNASLSVVEIKSPRLVLNDTCHLKS
      160      170      180      190      200

```

SEQ ID 9000 (GBS44) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 6; MW 27kDa), in Figure 168 (lane 8-10; MW 42kDa - thioredoxin fusion) and in Figure 238 (lane 7; MW 42kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 8; MW 52.4kDa).

Purified Thio-GBS44-His is shown in Figure 244, lanes 7 & 8.

Based on this analysis, it was predicted that these proteins and their epitopes would be useful antigens for vaccines or diagnostics.

50 Example 2223

A DNA sequence (GBSx2342) was identified in *S. agalactiae* <SEQ ID 6859> which encodes the amino acid sequence <SEQ ID 6860>. This protein is predicted to be d-alanyl-d-alanine carboxypeptidase. Analysis of this protein sequence reveals the following:

```

50     Possible site: 27
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----

```

-2506-

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 108/169 (63%), Positives = 139/169 (81%)

10 Query: 79 ELSPDVVFNITLYDKRITKQATQFLAARAIDSREHLISGYRSVAYQEKLFNSVYQEM 138
 E++PDV ++ +D RI + +FL AA+ IDS EHLISGYRSVAYQEL+L+N+Y+ QR
 Sbjct: 4 ESNPDVTDIDGVKVSRIENIRKFLAAQAQIDSSSEHLISGYRSVAYQEELYNNTYIAQEK 63

Query: 139 TSNPNLTRQAKRLVKTSQPAGASEHQTGLADMSTVDSLINESDPVSVQLKIAPIQY 198
 +NP+L++ +A+K V+TYSQF G+SEHQTGLA+DMSTVDSL+SD VV+++ IAP+YQ
 15 Sbjct: 64 ANNPSSLQBEAQKQVQTSQPFQSSSEHQTGLAIDMSTVDSLQSDANVAVKAAIAPKY 123

Query: 199 FVLRFPPDGKTASTGVGYEDMHYRYVGVSAKYMAKHILTLSEYITLLKE 247
 FVLRFPP+GK TG+ YEDMHYRYVGV+SAKYM KH LTLREY+ LKE
 20 Sbjct: 124 FVLRFPPGKDKATGIDYEDMHYRYVGVSAKYMTHKIDLTLEYLKLLKE 172

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6861> which encodes the amino acid sequence <SEQ ID 6862>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 25 INTEGRAL Likelihood = -9.66 Transmembrane 10 - 26 (3 - 29)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 118/173 (68%), Positives = 139/173 (80%)

35 Query: 74 ITKEMPELADINGISVDKRIBQATSDFLAAQAIDQLREHLISGYRSVDYQTELYQSYIK 133
 IT EM+P++ DI+G+ VD RI + T FLAAQ ID EHLISGYRSV YQ ELY +YI
 Sbjct: 1 ITAENPDVTDIDGVKVSRIENIRKFLAAQAQIDSSSEHLISGYRSVAYQEELYNNTYIA 60

40 Query: 134 KEMANDPTLTQEAALVQTSYQPPGASEHHTGLAIDMSTVDTLNASDPVSAKAVQKIA 193
 +E AN+P+L+QE A+ VQTSYQPPG+SEH TGLAIDMSTVD+IN SD +V V IAP
 Sbjct: 61 QSKANNPSSLQBEAQKQVQTSYQPPGSSSEHHTGLAIDMSTVDSLQSDANVAVKAAIAP 120

45 Query: 194 KYGFVLRFPPDGKTSYGVYEDMHYRYVGVSAKARYMAQHNLTLSEYITLLE 246
 YGFVLRFPPGKK +TG+DYEDMHYRYVGV SA+YM +H+LTLREY+ LKEK
 Sbjct: 121 KYGFVLRFPPGKDKATGIDYEDMHYRYVGVSAKARYMAQHNLTLSEYITLLE 173

An alignment of the GAS and GBS proteins is shown below.

Identities = 131/235 (55%), Positives = 172/235 (72%), Gaps = 3/235 (1%)
 50 Query: 15 LLAILCF--SLFALLKPNQSSSQKLARNEDIRKISSQKRKKLQLPVSQKDWNLIN 72
 LL ++ F L+ +KP ++ +Q L ++I++ +K ++ LP VS +DW L+LVN
 Sbjct: 12 LLIVIVFLGLGLYLFIPKEBSVPTIQ-LNKKIKQDKTRDLRALPKVSVDMELVLIN 70
 55 Query: 73 RDHHEELSPDVVFNITLYDKRITKQATQFLAARAIDSREHLISGYRSVAYQEKLFNS 132
 RDH +E+SP++ + I +DKRI + +FL AA+ID +EHLISGYRSV YQ +L+ S
 Sbjct: 71 RDHITKEMPELADINGISVDKRIBQATSDFLAAQAIDQLREHLISGYRSVDYQTELYQ 130
 60 Query: 133 YVTQEMTSNPNLTRQAKRLVKTSQPAGASEHQTGLADMSTVDSLINESDPVSVQLK 192
 Y+ +EM ++P L+ AR LV+TYSQF GASRH TGLA+DMSTVD+IN SDP V ++K
 Sbjct: 131 YIKEMANDPTLTQEAALVQTSYQPPGASEHHTGLAIDMSTVDTLNASDPVSAKAVQK 190
 Query: 193 IAPQYGFVLRFPPDGKTASTGVGYEDMHYRYVGVSAKYMAKHILTLSEYITLLKE 247

-2508-

The GBS18-His fusion product was purified (Figure 93A; see also Figure 189, lane 11) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 93B), FACS (Figure 93C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoinaccessible on GBS bacteria and that it is an effective protective immunogen.

5 Example 2224

A DNA sequence (GBSx2343) was identified in *S. agalactiae* <SEQ ID 6863> which encodes the amino acid sequence <SEQ ID 6864>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood =-12.58    Transmembrane    10 - 26 ( 3 - 29)

----- Final Results -----
bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6865> which encodes the amino acid sequence <SEQ ID 6866>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood =-11.83    Transmembrane    10 - 26 ( 4 - 33)

----- Final Results -----
bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

30 >GP:AAD00279 GB:U78599 putative N-acetyl-muramidase [Streptococcus mutans]
    Identities = 66/150 (44%), Positives = 97/150 (64%), Gaps = 5/150 (3%)

Query: 18 LLLIVCPLLSSORIASADKEVRVNYSQKQFITRMKEVKEPLAKYIGIRPSILIAQLLET 77
      LLI+ P+L+S +A A+K++ YS K+F+ ++ + L+K YG+R SI+I Q L++
35 Sbjct: 3 LLVILLPILASGLADANKKMPSPYSHKRFVKELIAPTAKLSKIYGVRSIIIGQAALDS 62

Query: 78 HDGNTLLASKYHNLFPSKKATPGQVAITLKS PKQTN---QNV--RYAIYKDDASAIRDYLR 132
      H G TLLASKYHNLFPS +A+PGQ A+ LKS + N Q V RY +Y+ ++ DY+
40 Sbjct: 63 HPGSTLLASKYHNLFPSLEASPGQAVRLKSHYKNGKRWQKVNTNRYLVYESWKESLVDYTA 122

Query: 133 MLRQKGEVDKRLRYNRLATKNGYKAPAKSLQ 162
      +L K DK LY + T GYK A++LQ
Sbjct: 123 ILHGNKINDKALYTTMTSSGYKTVARALQ 152

```

45 An alignment of the GAS and GBS proteins is shown below.

```

Identities = 67/190 (35%), Positives = 102/190 (53%), Gaps = 1/190 (0%)

Query: 1 MRKGFSLNLFIVVTFFIPFFFLPILNHRKQVDANSRQSVTTKEEFIQKIVPDADLQK 60
      MRKR F+ + F I+ PLI+ + A+ V Y+++FI K+ + + L K
50 Sbjct: 1 MRKRLKFPYFLTLACLLIVCPLLSSORIASADKEVRVNYSQKQFITRMKEVKEPLAK 60

Query: 61 SYGIRPSFIITAAALDSDFGEKILLANKYHNLFGLLAEPTGPTSLITLNDSTGKKQKQKPTH 120
      YGIRPS +IAQ L++ G+ +LA+KYHNLF A PG +ITL S Q ++
55 Sbjct: 61 YYGIRPSILIAQLLETIDGKNTLLASKYHNLFPSKKATPGQVAITLK-SPKQTNQNRYAI 119

Query: 121 YKSWKYSMYDYLAHKISGATGKDDSYTIMVSVNPKTLVQKLDGSGFNDKKYAKNMTEI 180
      YK ++ DYL ++ G K Y + + K K + LQ DK YA++ ++

```

-2509-

Subject: 120 YKDDASAIKDYILFRLQSGKEVDKRLYRNLAIEGKYKAPKSLQKYLHYTDTKYARRLIQV 179

Query: 181 IDLYDLTRYD 190
1+ DEL YD

Subject: 180 IESNDLTRYD 189

SEQ ID 6864 (GBS246) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 7; MW 24.6kDa).

GBS246d was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lanes 14 & 15; MW 21kDa) and in Figure 183 (lane 4; MW 21kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 12; MW 46kDa). Purified GBS246d-GST is shown in Figure 243, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2225

A DNA sequence (GBSx2344) was identified in *S. agalactiae* <SEQ ID 6867> which encodes the amino acid sequence <SEQ ID 6868>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2541(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45610 GB:U78296 repressor of class I heat shock gene
expression HrcA [Streptococcus mutans]
Identities = 227/345 (65%), Positives = 287/345 (82%), Gaps = 1/345 (0%)

Query: 17 VITQRNDIINLIVELPTCTHEPVGSKALQRTIDSSATIRHMDAIEKLGILLEKATSS 76
+ITQRQ DIINLIVELPT+THEP+GSK LQ +I 88 ATIRHMDA LKLGILLEKA T
Subject: 1 MITQRQDIINLIVELPTCTHEP IGSKTQNSIASSRATIRHMDAIEKLGILLEKATPP 60

Query: 77 GRM-PSFAGFKYFVHSLRLDSIDSDQIYHVLIKAPDFEAPKLEMDQKASHILSEMTGTT 135
+ P +YFVHSL DS+DEQD+Y VIKAPDFEAP+L D+LQ+AS +L+ +RGTT
Subject: 61 AVVCFVKKAIKRYFVHSLNPDLSLDQDVYQVIKAPDFEAPKLGDLQASDVLANLIGTT 120

Query: 136 SVLLDVEPARQLCGFVVQLSNHDAVMTLDESQKPVTVQPAIRPNFLTRDLAPKALV 195
++LLDVEP +QRIT FD+V+LSNHDAV+TLDE+ PVTQVPAIP+NEL DL+ I
Subject: 121 ALILDVEPKQRILTTFDVLKLSNHDAVMTLDESQKPVTVQPAIRPNFLTRDLAPKALV 180

Query: 196 EERLDGSGVMDIHKLRTETPCIVQKQYFVTIMVVLQLEFDYVPSLEPTEVFFVAGKVNST 255
ER L+ +V+DIHY+LRT EPCI+QKFP TDMVL LED+P+ +F E VP+GK+ +L
Subject: 181 EERFLAQTVLIDHYRLRTEPPQIIQYFPRTIMVLDLFDHIFNDIPQEEVPIGSKIKTYE 240

Query: 256 YSGLSTYQFLDNEQQVAISRLQSLKRGEMASQVADSQRAALADVSVLTHKFLIPYRGFG 315
++ L TYQFL+N Q VA+ +RQSL R E+ VQVADS+E +LAD+V++ KFLIPYRGFG
Subject: 241 FAGLSTYQFLDNEQLSVAELTRQSLPEDELRHVQVADSKEKSLADLVISQKFLIPYRGFG 300

Query: 316 LLSLIGPIDMVRVRSVLANIIGKVLAAKLIDYHYRLNSNHYEVH 360
+L+ +IG+D+DY+R++SL+N+I +VLA KLGD+YRYLNSNHYEVH
Subject: 301 ILIVIGPVDLDYQRTISLIVLSRVLAVKLDGVPYRYLNSNHYEVH 345

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6869> which encodes the amino acid sequence <SEQ ID 6870>. Analysis of this protein sequence reveals the following:

-2510-

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0695 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 341/344 (99%), Positives = 343/344 (99%)

Query: 17 VITQRQNDIILNLI VLFPTQTHFVPGSKALQRTIDSSATIRNDMAKLEKGLLEKHAHTSS 76
 VITQRQNDIILNLI VLFPTQTHFVPGSKALQRTIDSSATIRNDMAKLEKGLLEKHAHTSS
 Sbjct: 1 VITQRQNDIILNLI VLFPTQTHFVPGSKALQRTIDSSATIRNDMAKLEKGLLEKHAHTSS 60

15 Query: 77 GRMPSPAGFKYFVHSLRLDSIDBQDIYHVIKAFDFEAFKLEDMQKASHILSMTGYTS 136
 GRMPSPAGFKYFVHSLRLDSIDBQDIYHVIKAFDFEAFKLEDMQKASHILSMTGYTS
 Sbjct: 61 GRMPSPAGFKYFVHSLRLDSIDBQDIYHVIKAFDFEAFKLEDMQKASHILSMTGYTS 120

20 Query: 137 VILDVEPARQRLTGFDVVQLSNHDAVMTLDSKPVTVQFAIPRNFLTRDLIAFKAIVE 196
 VILDVEPARQRLTGFDVVQLSNHDAVMTLDSKPVTVQFAIPRNFLTRDLIAFKAIVE
 Sbjct: 121 VILDVEPARQRLTGFDVVQLSNHDAVMTLDSKPVTVQFAIPRNFLTRDLIAFKAIVE 180

25 Query: 197 BRLLDGSVMDIHYKLRTIPIQIVQKYFVTNDVLQLFDFYVPSFLFLETVPVAGK/NSLTY 256
 BRLLD SV+DIHYKLRTIPIQIVQKYFVTNDVLQLFDFYVPSFLFLETVPVAGK/NSLTY
 Sbjct: 181 BRLLDNSVIDIHYKLRTIPIQIVQKYFVTNDVLQLFDFYVPSFLFLETVPVAGK/NSLTY 240

 Query: 257 SDLSTYQFLNDBQQVAISLRQSLKGBGMSVQVADSQRAALADSVLTHKFLIPYRGGL 316
 SDLSTYQFLNDBQQVAISLRQSLKGBGMSVQVADSQRAALADSVLTHKFLIPYRGGL
 Sbjct: 241 SDLSTYQFLNDBQQVAISLRQSLKGBGMSVQVADSQRAALADSVLTHKFLIPYRGGL 300

30 Query: 317 LSLIGPIDMYRRSVSLVNIIGKVLAALKGDYYRYLNSNHYEVH 360
 LSLIGPIDMYRRSVSLVNIIGKVLAALKGDYYRYLNSNHYEVH
 Sbjct: 301 LSLIGPIDMYRRSVSLVNIIGKVLAALKGDYYRYLNSNHYEVH 344

35

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2226

A DNA sequence (GBSx2345) was identified in *S.agalactiae* <SEQ ID 6871> which encodes the amino acid sequence <SEQ ID 6872>. This protein is predicted to be grpe protein (grpE). Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.5138 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database.

>GP:AA045611 GB:U78296 GrpE [Streptococcus mutans]
 Identities = 130/180 (72%), Positives = 151/180 (83%), Gaps = 3/180 (1%)

55 Query: 14 VSEIHKDDLLQEEVENTE--TESTVEEVIESIPKSELILANRADEPENKYLRAPHAEM- 70
 +8+ + KK++ +EEVATE TE+VEEV EE E EL+ A ERA++PENKYLRAPHAEM
 Sbjct: 1 MSKKDKKREYKEEVATEPTTSEVEEVAESTSRNKLQALSRARDPENKYLRAPHAEM 60

 Query: 71 QNIQRSSSEERQQLQRYRSQDLAKAILPGLNLERALAVESLTDDVKGLEMTDRSLQA 130
 + + + QRYRSQDL KAILPGLNLERALAVESLTDDVKGLEMTDRSLQA
 Sbjct: 61 KTPSVALKMSDKVQRYRSQDLKAILPGLNLERALAVESLTDDVKGLEMTDRSLQA 120

60

-2511-

Query: 131 LKESGVVEEVDSPDHNPHMAVQTLPADDEHPADSLAEVFGKGYKLHERLLRPAMVVYN 190
 LKESGVVEEVE++FD N HPAVQTL ADD+HDAASIA+V QGGY+LHERLLRPAMVVYN
 Sbjct: 121 LKESGVVEELENFDANLEHMAVQTLADDDHPADSLAQVHQGYQLHERLLRPAMVVYN 180

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6874> which encodes the amino acid sequence <SEQ ID 6874>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 189/190 (99%), Positives = 189/190 (99%)

Query: 1 MAVFNKLFKRRHSVSEIKKDDLQEEVEATETETVEEVIEE PEKSELELANERADEF 60
 MAVFNKLFKRRHSVSEIKKDDLQEEVEATETETVEEVIEE PEKSELELANERADEF
 Sbjct: 1 MAVFNKLFKRRHSVSEIKKDDLQEEVEATETETVEEVIEE PEKSELELANERADEF 60
 Query: 61 NKYLAAHAEMQNIQRSSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120
 NKYLAAHAEMQNIQRSSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL
 Sbjct: 61 NKYLAAHAEMQNIQRSSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120
 Query: 121 EMTRDSLIQALKEGVEEVDSPDHNPHMAVQTLPADDEHPADSLAEVFGKGYKLHERL 180
 EMTRDSLIQALKEGVEEVDSPDHNPHMAVQTLPADDEHPADSLAEVFGKGYKLHERL
 Sbjct: 121 EMTRDSLIQALKEGVEEVDSPDHNPHMAVQTLPADDEHPADSLAEVFGKGYKLHERL 180
 Query: 181 LRPMAMVVYN 190
 LRPMAMVVYN
 Sbjct: 181 LRPMAMVVYN 190

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2227

A DNA sequence (GBSx2346) was identified in *S.agalactiae* <SEQ ID 6875> which encodes the amino acid sequence <SEQ ID 6876>. This protein is predicted to be heat shock protein 70 (dnaK). Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0996(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6877> which encodes the amino acid sequence <SEQ ID 6878>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0996(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2512-

An alignment of the GAS and GBS proteins is shown below.

Identities = 594/609 (97%), Positives = 603/609 (98%), Gaps = 1/609 (0%)	
5	Query: 1 MSKIIIGIDLGTNSAVAVLEGTESKIIANPEGNRTTTPSVVSFKNGRIIVGDAARKQAVTN 60
	Sbjct: 1 MSKIIIGIDLGTNSAVAVLEGTESKIIANPEGNRTTTPSVVSFKNGRIIVGDAARKQAVTN 60
10	Query: 61 PFTVISIKSKMWTSEKVSANGKEYTTPQRIISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120
	Sbjct: 61 PFTVISIKSKMWTSEKVSANGKEYTTPQRIISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120
15	Query: 121 NDAQRQATKDAGKIAGLEVERIVNEPTAALAYGMOKTDKDEKILVFDLGGTDFVDSILE 180
	Sbjct: 121 NDAQRQATKDAGKIAGLEVERIVNEPTAALAYGMOKTDKDEKILVFDLGGTDFVDSILE 180
20	Query: 181 LGDGVFDVLATAGDNKLGSDDFDQKIIDFLV EFKENGIDLSQDKMALQLKDAAEKAK 240
	Sbjct: 181 LGDGVFDVLATAGDNKLGSDDFDQKIIDFLV EFKENGIDLSQDKMALQLKDAAEKAK 240
25	Query: 241 KDLGVTQTQISLFPFITAGSAGPLHLEMSLSRAKFDLTRLVETKTFVRQALSAGLS 300
	Sbjct: 241 KDLGVTQTQISLFPFITAGSAGPLHLEMSLSRAKFDLTRLVETKTFVRQALSAGLS 300
30	Query: 301 LSEIDEVILVGGSTRIPAVVEAKVAETGKEPKNSVNFDEVVAMGAIIQGVITGDVKDV 360
	Sbjct: 301 LSEIDEVILVGGSTRIPAVVEAKVAETGKEPKNSVNFDEVVAMGAIIQGVITGDVKDV 360
35	Query: 361 LLDVTPLSLGIETMGVFTKLIDRNTTIFTSKSVFTTAADNQPAVDIHVLQGERPMAAD 420
	Sbjct: 361 LLDVTPLSLGIETMGVFTKLIDRNTTIFTSKSVFTTAADNQPAVDIHVLQGERPMAAD 420
40	Query: 421 NKTGLRQFLTDIPAAFGPIQIEVTFIDKNGIVSVKAKDLSTQKQSHIVT+SM CL++E 480
	Sbjct: 421 NKTGLRQFLTDIPAAFGPIQIEVTFIDKNGIVSVKAKDLSTQKQSHIVT+SM CL++E 480
45	Query: 481 EIDRMMKDAEANAADAKRKEEVDLIGNEVDQAI FATEKTIKETBGKGFDFTERDAQAQALD 540
	Sbjct: 481 EIDRMMKDAEANAADAKRKEEVDLIGNEVDQAI FATEKTIKETBGKGFDFTERDAQAQALD 540
50	Query: 541 ELKAAQESGNLDMKAKLEALNEKAQALAVKYEQAAAAQAAQAGAEQAQADSSSGDD 600
	Sbjct: 541 ELKAAQESGNLDMKAKLEALNEKAQALAVMYEQAAAAQAAQAGAEQAQANDSAN+KDD 599
55	Query: 601 VVDGEPTK 609
	Sbjct: 600 VVDGEPTK 608

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 2228

A DNA sequence (GBSx2347) was identified in *S. agalactiae* <SEQ ID 6879> which encodes the amino acid sequence <SEQ ID 6880>. This protein is predicted to be *Streptococcus pneumoniae* DnaJ protein homologue (dnaJ). Analysis of this protein sequence reveals the following:

55	Possible site: 18
	>>> Seems to have no N-terminal signal sequence
	----- Final Results -----
	bacterial cytoplasm --- Certainty=0.4180 (Affirmative) < succ>
	bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
60	bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-2514-

Query: 4 KVEIRSYLIASIQNGKLAGDRLPSIRQLANQSNCKDTQVQVLMELRFQNYIYAK 63
K + I ++ I + G + LPSIRQL Q + C + KDIVQ + + EL + N IYA + S
Sbjct: 3 KYQVITQILITGIEHRFKRGEKLPISRLQKQYHCSDIVQKAMELKYQNKIYAVEKS 62

5 Query: 64 GYTVFDSHQEVEBGVSLPNSIEIANIAYDDPRICLANETLIGREDYLPYHYHQEGLLD 123
GYT + + + + + ++ I Y + DR + CL E + LIGRE + YLPYHY + QEGEL + L
Sbjct: 63 GYTILEDREDPQ-DHTCRAQSYRLSRITTYEDPRICLKSRLIGRENYLPYHYHQEGELALI 121

10 Query: 124 KAVAKLMEETGVTVPLDDIVITAGTQQAFLITQVTFPNRKSRLVLEPTYPRIEIL 183
+ V L + + VY D + VITAG + QQA + ILTQ + K + + LIR PTY RMEIL +
Sbjct: 122 SSVQSLMDYHVTYTKDQLVITAGSQALYILTQMETLAGKTEILHMTYRMIELIRH 181

15 Query: 184 QNLPEYETISRGTHIGDIPQRELEIPQOSIKFFVYPRMNPGLTSSYNPVMKRLIEMAEK 243
Q + PY + TI R CID + LE IPQT IKFFY IPR + HNPGL + Y + + + + + +
Sbjct: 182 CGIPKQTIERNLDGIDLESLIPQTKIKFFYTPRLHNPGLSTDIATKTAIVKLAKQ 241

20 Query: 244 YDVYIYEDDYMDSDFASQ--PLHYDTHGRVYILKSPSKAIFPALRLAALCLPQALKST 301
YDVYI + EDDY + DF S PLHY DT RVYI + KSP + +FPALR + AI LP L F
Sbjct: 242 YDVYIIEDDYADFDSSHSLPLHYDTDRVYIYKSTPTFLPALRIGALSNQLRDIF 301

25 Query: 302 MAYKKLMDYDTNLILQKALALYIENGLYAKNSQYIAKYQKDLANSKSLADHP-NLPSY 360
+ K L + DYTNLIL + QKAL + LYI + NG + + A + N + Q + L + Y K L + N + P Y
Sbjct: 302 IKHKSILIDYTNLIMQKALSLYIDKGMFAFNTQHLHYIYHQAQNKIKDCLKLYALNIP-Y 360

30 Query: 361 SLHSDGVLPDCKS 373
+ SV F SK
Sbjct: 361 RYFKGSVTFQLSK, 373

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6885> which encodes the amino acid sequence <SEQ ID 6886>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3043 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 176/382 (46%), Positives = 255/382 (66%), Gaps = 7/382 (1%)

Query: 1 MVTKEIRSYLIASIQNGKLAGDRLPSIRQLANQSNCKDTQVQVLMELRFQNYIYAK 60
M TK + I S + IQ + L GD + LPSIR L + C + KDIVQ R L + EL + + IYA
Sbjct: 1 MTKYQTIENIBQDIQKRLKKGDKLPISRLSKVYCSKDTQVQALLKLYRHLIYAV 60

45 Query: 61 PRSGYTVFDSHQEVEBGVSLPNSIEIANIAYDDPRICLANETLIGREDYLPYHYHQEGLL 120
P + SGYTV + + ++ L + N + A + Y + DR + CL E + LIGRE + YLPYHY + QEGEL + L
Sbjct: 61 PRSGYTVL-GNVSMFDNVLNLSLEDYNNMAYEDPRICLANEALSAXDKYLPYHYHTKSL 119

50 Query: 121 DLSKAVAKLMEETGVTVPLDDIVITAGTQQAFLITQVTFPNRKSRLVLEPTYPRIEIL 180
+ L + A + + E VY D + IT + GTQQA + IL + Q + PPN + L + E + PTY RM +
Sbjct: 120 ELREALLYLAENSVSNKQDLITSGTQQAFLITSCMPPTGKTIILEKPTYHMEAI 179

55 Query: 181 IKTNLPEYETISRGTHIGDIPQRELEIPQOSIKFFVYPRMNPGLTSSYNPVMKRLIEMAEK 240
+ LPY + TISR + GAD + LE + FQT IKFFY I R + + PLG SY + E + + + +
Sbjct: 180 VAQLGLPYQTISRHFNGLDLELLESLPQTGDIKFFYTPRSHNPGLSYSTKEKRAIVRL 239

60 Query: 241 AERYDVYIYEDDYMDSDF--SQSLPHYDTHGRVYILKSPSKAIFPALRLAALCLPQALK 298
A + + Y VYI + EDDY + DF + P + HYDTH R + IYKSPS + +FPALR + A + LP LK
Sbjct: 240 AQRQYVYILKDDYGLDFVILKRPPIHYDTHRIIYKSPSVPFALRIGALVLPGLK 299

Query: 299 STPMAYKKLMDYDTNLILQKALALYIENGLYAKNSQYIAKYQKDLANSKSLADHPNLP 358
P + K L + D DTNLIL + QKALALY + ENG + + KN + + K RY K + + +
Sbjct: 300 PHFLTQKSLIDLDYTNLIMQKALALYENGMFQKNLRPIK-RYLKQREKRALFLQ-NCP 357

65

-2515-

Query: 359 S--YSLHSDSVLFDCSKLDNFK 378
 Y L ++ D + D+++
 Sbjct: 358 DHYQLTPHIVIDYTTSDSYR 379

- 5 SEQ ID 6884 (GBS423) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 7; MW 49.3kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 2; MW 74kDa).

GBS423-GST was purified as shown in Figure 219, lane 2-3.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 2230

A DNA sequence (GBSx2349) was identified in *S. agalactiae* <SEQ ID 6887> which encodes the amino acid sequence <SEQ ID 6888>. This protein is predicted to be pseudouridylyl synthase I (truA). Analysis of this protein sequence reveals the following:

- 15 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.3265 (Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:BA003886 GB:AP001507 tRNA pseudouridine synthase A
 (pseudouridylyl synthase I) [Bacillus halodurans]
 Identities = 105/240 (43%), Positives = 147/240 (60%), Gaps = 2/240 (0%)
- Query: 1 MTRYKQTSYDGSAPSGPQRPNCRTVQEEIERLTKRINSQNDVIIHGAQRTDVGWHAYG 60
 M R +++YDG+ F+G+Q QPN RTVQ E+E LK ++ G + + +GRTD GVHA G
 30 Sbjct: 1 MKRIGLKVAYDGTDFAGYQIQPNERTVQSELVSLKNIHKGMSIRVTASGRITDGVHAG 60
- Query: 61 QVHFPLQARDVEKIRFGIDTQCDDIDIVKVBQVSDDFHCYRDKHINKTYEFLVDIGRP 120
 Q++HFD + V++ L++Q P DI +++ V DFH RY K Y + V
 35 Sbjct: 61 QIVHFDTSLSFPVDRMPIALNSQLPADI CVLEADVADPHARYSAKTKYRYRVLTSAQ 120
- Query: 121 KNPMRNYATHYPYPVILMLQEAIKDLVGTHDFTGFTASGTSVENKVRTIFDAKIQFEA 180
 + RNY H YP+ +R MQ A L+GTHDF+ F A+ VE+KVRTI D + E
 35 Sbjct: 121 ADVFRNRYTHYRVPLDVEAMQRAAVQLGTHDFSSFCAAKAEVBEKVRTIEDVALMREG 180
- Query: 181 SKNLLIFTFTNGFLYKQVRNMGVTLKIGNGRMPISQIKTILQAKNRDLAQPTAAGNL 240
 + LIF+ NGFLY VR +VGTL+IG G+ ++ IL A++R+ AG TA G+GL
 40 Sbjct: 181 DE--LIFSIRNGFLYNVRIIVGTGLEIGAGKRSAREVAKILAAERSRAGKTAAGHGL 238

- A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6889> which encodes the amino acid
 45 sequence <SEQ ID 6890>. Analysis of this protein sequence reveals the following:

- Possible site: 58
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.2558 (Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2516-

Identities = 184/249 (73%), Positives = 214/249 (85%)

Query: 1 MTRYKATISYDGSASFSGFCRQPHCTVQRRIERTLKRINSQNDVLIHGAGRTDUGVHAYG 60
 M RYKA ISYDG+ PSGFQRO + RTVOREIE+TL +L+G +LIHGARTD GVHAYG
 5 Sbjct: 1 MVRKATISYDGTLPFGFQQRHLRTVQSEIEKTLKLNNGTKLIHAGARTDAGVHAYG 60

Query: 61 QVIHFOLFQARDVEKLRFGLDTQCDDIDIVVEQVSDDFHCVRDKHKIKTYEPLVDIGRP 120
 QVIHFOLFQ ++VEKLRF LDTQ P+DID+V +B+V+DDPHCRY KH+KTYEPLVD GRP
 10 Sbjct: 61 QVIHFOLFQEQEVEKLRFALDTCTPDDIDVNNIKVADDFHCYKQHLKTYEPLVDIGRP 120

Query: 121 KNPMGRNVAHYTHYPYVILIMQEAIKDLVGHDPGTGPTASGTSVENKVRTIPDAKIQFEA 180
 KNPMGR+Y THYPY + I+IMQEA I LVOGHDPGTGA+GTSV+KVRTI A + +
 Sbjct: 121 KNPMGRHYTHYPYVITLNIKMQEAINGLVGHDPGTGAAGTSVQNKVRTITKATVSRDE 180

Query: 181 SKNLLIFTFNGFLYKQVRNMGVTLKLGNGRMPISOIKTILQANRDLAGPTAAGNGL 240
 + L+FTF+GNGFLYKQVRNMGVTLKLGNG+MP+ Q+K IL +KNR LAGPT +GNGL
 15 Sbjct: 181 KTDPLVTFPSGNGFLYKQVRNMGVTLKLGNGQMPVEQVKVILSKNRQLAGPTISGNGL 240

Query: 241 YLKEIIVED 249
 YLKEI YE+
 20 Sbjct: 241 YLKEICYEN 249

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2231

A DNA sequence (GBSx2350) was identified in *S. galactiae* <SEQ ID 6891> which encodes the amino acid sequence <SEQ ID 6892>. This protein is predicted to be phosphomethylpyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2051 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB15828 GB: Z99123 phosphomethylpyrimidine kinase [Bacillus subtilis]
 Identities = 95/253 (37%), Positives = 150/253 (58%), Gaps = 13/253 (5%)

40 Query: 1 MKTENVLAISGNDIPSGGGLHADLATYVNNKLHGFVAVTCLTAMSDDG---FEVIPERAS 57
 M L I+G+D G G+ ADL T+ ++G A+T + AM +V PI+
 Sbjct: 1 MSMEKALITAGSDSSGAGIQADLKTPOEKNVYGMTALTIVIMDDPNNEWNHQVFFIDTD 60

45 Query: 58 ILKQQLBSLKD-VEFGSIKIGLLPNVETAQVVLRFVKSQKQCPVVLDPVLVCKENHDL-- 114
 ++ QL ++ D + ++K G+LP V+ ++ +K KQ VV+DPV+VCK +++
 Sbjct: 61 TIRAQALATITDIGVDAMKTNLPTVDIELAAKTIKQKLNVDVDPVWCKGANVSLY 120

Query: 115 --EVSQLRQQLIAFFPYADVITPNLVEAQLLTGLS-IENLDQMKIAEKIYDMGAKHVVI 171
 LREQQL P A VITPNL EA L+G+ ++D M AA+K+ ++GA++VVI
 50 Sbjct: 121 PEHAQALREQLA---PLATVITPNLFEASQSGNDELKTVDOMIEAAKIHAGAQYVVI 177

Query: 172 KGNRLNASEATDLYYDGERFETVFPVVDANNT-GAGCTFASIASQLAMGNVEDAVK 230
 GG +L B+A D+ YDGE E ++D T GAGCTF+ +++ ++LA G V++A+
 55 Sbjct: 178 TGGCKLHKRKAVDVLYDGETAEVLESEMIDTPYTHGAGCTFPAAVTAEIAKGAQVKEIAY 237

Query: 231 MSKGFVYQAIKAS 243
 +K P+ AIK S
 Sbjct: 238 AAKSFITAAIKES 250

60

-2517-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4407> which encodes the amino acid sequence <SEQ ID 4408>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2029 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 135/252 (53%), Positives = 174/252 (68%)

Query: 1 MKTRNVLAISGNDIFSGGGLHADLATYVVKLHGPFVAVTCLTAMSDRGFEVPIEASILK 60
MKT ++ ISGNDI SGGGL+ADLATY+ L FVAVTCLT S++GF+ P+ I +
Sbjct: 1 MKTDVITVITSGNDISGGGLADLATYIIRYDLQAFVAVTCLTTRSEEGFSLPFAKEIFR 60

Query: 61 QQLSESLKIVSFGSIKILGLLPVNETAQVLEFVKSKQECFVVLDPVLVCKENHDLVLSQLR 120
Q S + +IK+GLLEN E ++VL+P+K PVVLDPVL CHE D+++ LR
Sbjct: 61 DQINSFTNAPISAIKIGLLENACMEIIVLDFIKGHLGIPVVLDPVLVCKEIDVVKIVFLR 120

Query: 121 EQLIAFFPYADVITPHLVEAQLTGLSIENLDMKIAAEKLYDMGAKHVYIKGGNRLNAB 180
+++ FY V+PHLVEAQLT+ I +L M+ AA+ Y +GAK VVIKGGNR + +
Sbjct: 121 QEILQLLPYVTVVTPHLVEAQLSQKEIVSLKDMQEAAYFYQLGAKQVYIKGGNRFSQK 180

Query: 181 EATDLYYDGERFETTVFVVDDNTGAGCTPAGSTASQLAMGRNVEDAVINSGKGFVYQAI 240
+A DL+YDG+ T PV++ NN GAGCTPAGSTASQL K +AVK SK VYQAI
Sbjct: 181 KALDLYYDGKEIVTLECFVLEKNNIGAGCTPAGSTASQLVKKKTFLEAVNKGKELVYQAI 240

Query: 241 KASDKYGVQVHF 252
SD+YGV Q +
Sbjct: 241 LQSDRYGVQKSY 252
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2232

A DNA sequence (GBSx2351) was identified in *S.agalactiae* <SEQ ID 6893> which encodes the amino acid sequence <SEQ ID 6894>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -6.05 Transmembrane 97 - 113 ( 96 - 119)
INTEGRAL Likelihood = -0.22 Transmembrane 54 - 70 ( 54 - 70)

----- Final Results -----
bacterial membrane --- Certainty=0.3421 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA30952 GB:AP000007 202aa long hypothetical protein [Pyrococcus
horikoshii]
Identities = 48/148 (32%), Positives = 78/148 (52%), Gaps = 9/148 (6%)

Query: 10 VOLAIVTAISIVLGMFISIPITPGLTLLEDGIFFAAFYRKKKBAVVGAGLGLDLK 69
V +A+VTA++V+ I IP G+L D I + PG G G + DLL
Sbjct: 49 VMAALATATWVIR--IPIPASQGYINFGDIMMLTSLVPLFGVGFAGVGSFADLL- 105

Query: 70 GYPNMFPSLLIHGTQVYLAGLGR-----RRLGLISATLVVVLGYATASGLMYGNGR 123
GYP+N E+L+I GT+G + G + + LLG + VMV+GY + ++YG
```

-2518-

Sbjct: 106 GYPSWALFTLVINGTSGIIVGYPSKGRANYGKILLGTVLGSSVMVIGYVSVAIVLYGPGAG 165

Query: 124 VLFDIPGNIMQNMGMVVGWFAINKSLER 151

++ +I+Q +G+V+G L L++

5 Sbjct: 166 AIGELYNDIVQAVSGIVIGGSLGYLKK 193

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6895> which encodes the amino acid sequence <SEQ ID 6896>. Analysis of this protein sequence reveals the following:

Possible site: 54

10

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -4.62 Transmembrane 98 - 114 (97 - 119)

INTEGRAL Likelihood = -0.00 Transmembrane 135 - 151 (135 - 151)

15

----- Final Results -----

bacterial membrane --- Certainty=0.2848 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20

The protein has homology with the following sequences in the databases:

>GP:CB49310 GB:A248284 hypothetical protein [Pyrococcus abyssi]

Identities = 42/145 (28%), Positives = 73/145 (49%), Gaps = 10/145 (6%)

25

Query: 7 RQNSLTGILITLVVVLGRFVMLPTPT--GFLTLDDAGIYAVSPFQSAQGAIVGSLGFL 64

R +++++ ALV + + + P P G+L D I V+ PG G G3+ +

Sbjct: 39 RTVAISAVAAALVTAMTMVIRIPASQGYINFGDINIMLVAVLEPGFLVGGPAGGVSAT 98

30

Query: 65 IDLVAGYGPWMFSLIAHSVQGYFAGWRCR-----KRWLGUVIGSFMIFWYFLGSLML 118

DL+ GYP W +LI +G G+ + K +G V+G FIM+ Y S +L

Sbjct: 99 ADLI-GYPSWALFTLIIKSGELVVGYPFSGKGRPNYSKILIGTVLGGFINVLGYVSVSVL 157

Query: 119 GYGLSGSLAGIWMVQNTLGLFGL 143

YG +G+++ ++ +Q G+ +G

35

Sbjct: 158 -YGPAGAISELYNDIVQAVSGIVIG 181

An alignment of the GAS and GBS proteins is shown below.

Identities = 77/155 (49%), Positives = 106/155 (67%), Gaps = 1/155 (0%)

40

Query: 1 MKNEKTSQCLQALIVTAISVLGWFISIPTPTGFLTLDDAGIFFAFYFGKKGGAUVGAL 60

M+ K C+ I+TA+ +VLG F+ +PTPTGFLTLDDAGI+ +F FG +GA+VG L

Sbjct: 1 NGNSKIRQMSLTGILITLVVVLGRFVMLPTPTGFLTLDDAGIYAVSPFQSAQGAIVGGL 60

45

Query: 61 AGFLIDLLEKGYPNWVMSFLLINGTQGYLAGLQRRRLGLGLISATLVNVLGYTAISGLN-Y 119

+GFLIDL+ GYP WMF SL+ H QGY AG GR+R LG++ +M+ Y + S ++ Y

Sbjct: 61 SGFLIDLVAGYGPWMFSLIAHSVQGYFAGWRCRKRKRWLGUVIGSFMIFWYFLGSLMLG 120

Query: 120 GWCVAFLDIPGNIMQNMGMVVGWFAINKSLERVKK 154

G L -I GR+MQN +G+ VGF + K++ R KK

50

Sbjct: 121 GLSGSLAGIWMVQNTLGLFVGFIIKALLRQKK 155

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2233

55

A DNA sequence (GBSx2352) was identified in *S.agalactiae* <SEQ ID 6897> which encodes the amino acid sequence <SEQ ID 6898>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2519-

bacterial cytoplasm --- Certainty=0.0881(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CB15708 GB:Z99122 alternate gene name: ipc-33d [Bacillus subtilis]
 Identities = 91/176 (51%), Positives = 115/176 (64%)

10 Query: 6 NKLQETKAIVVDIERSALKKGQIFVLGLSSSEVSQGLIKNSSEIGEIIIVEVILKEL 65
 N+LQ K + + + LK+ Q+FVLG S+SEV+G IG + S + I E I + +
 Sbjct: 2 NELKQTKWIKLSEFQDAELKQDQLFVLGCTSEVAGSRIGTSGSVDIARSISYGLAELR 61

Query: 66 HSRGIYLVAGQCEHVNRALVVERALAEARQQLEVVNVVFNHAGGSGQVAAPKMTSPFEV 125
 GI+LA Q CEH+NRALVVERA+ +L V+ VP AGG+ AFK M SPV V
 15 Sbjct: 62 EKTGIHLAQCCERHARALVVERATAKIPRIPTVSAPVFPKAGGAMASYAFKMKSPVLV 121

Query: 126 EEIVAHAGIDIGDTSIGMHKRVQVPLIPISRELGAHVITALASPKLIGGARAGY 181
 E I A AGIDIGDT IGMH+K V VP+ LG AHVT +RPKLIGG RA Y
 20 Sbjct: 122 ETIQADAGIDIGDTFIMHLPKPAVVRVVSQNLGSAHVITALARTPKLIGGVRAVY 177

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6899> which encodes the amino acid sequence <SEQ ID 6900>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2166(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.
 Identities = 132/183 (72%), Positives = 161/183 (87%)

35 Query: 6 NKLQETKAIVVDIERSALKKGQIFVLGLSSSEVSQGLIKNSSEIGEIIIVEVILKEL 65
 N L+++T+ IV+D++ERSA++ G +FVLGLSSSE+ G IKG SS E+G+I+VEV+L EL
 Sbjct: 3 NNLEKQTRIVDIVERSAIQGNLFLVLGLSSSEILGSRIGKQSSLEVVQIVVEVILDEL 62

Query: 66 HSRGIYLVAGQCEHVNRALVVERALAEARQQLEVVNVVFNHAGGSGQVAAPKMTSPFEV 125
 + RG++LAVQGCEHVNRALVVR +AE +QLE+VNVVFNHAGGS Q+ARF+LM+ PVEV
 40 Sbjct: 63 NKRGVHLVAGQCEHVNRALVVERIVAESKQLEIVNVVFNHAGGSAQAARFQLMSDPFEV 122

Query: 126 EEIVAHAGIDIGDTSIGMHKRVQVPLIPISRELGAHVITALASPKLIGGARAGYSDP 185
 EE++AHAG+DIGDT+IGMHKRVQ+PLIP RELOGAHVITALASPKLIGGARA Y D
 45 Sbjct: 123 EEIVAHAGLDIGDTAIGMHKRVQVPLIPQRELOGAHVITALASPKLIGGARADYNQDI 162

Query: 186 IRK 188
 IRK
 Sbjct: 183 IRK 185

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2234

A DNA sequence (GBSx2353) was identified in *S.agalactiae* <SEQ ID 6901> which encodes the amino acid sequence <SEQ ID 6902>. Analysis of this protein sequence reveals the following:

55 Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -11.25 Transmembrane 21 - 37 (13 - 46)
 INTEGRAL Likelihood = -4.30 Transmembrane 78 - 94 (76 - 113)
 INTEGRAL Likelihood = -2.07 Transmembrane 96 - 112 (95 - 113)

-2520-

----- Final Results -----

5 bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 105/261 (40%), Positives = 150/261 (57%), Gaps = 2/261 (0%)

10 Query: 12 NVEEVLTFPTFKLIS--ILLIIAIPVIVRQVINYLFKFTVNRSLAPSQRQVARKTLAKL 69
 N+ P T +I+ +L+ +IAF+IVR + + + R ++ R TL KL
 Sbjct: 7 NITSGAFLASTFIILAGVILVAVIAFLIVRAIGKRIISNSFARMAKNNQLSSGRVVTLEKL 66

15 Query: 70 SHIVLYNTLYFFLFYWLISLGVPISSLLAGAGIAGVAILGAAGQGLSDVNVNGFTILEN 129
 S N +YTL F +L+I G+ S+L+AGAGI G+AI G GAQG +SD+V GPFILE
 Sbjct: 67 SINAFSYTLMFIFATILLITPGLNPSALAGAGIAGVILGAIGPGAQGLVSDIVTGVFTILEK 126

20 Query: 130 QFDVGDIIINVGVSGVTINVGIRITQIHDFDGLHFIIPNRNITIVSNKSRNMRALQIDIP 189
 Q DVG D + G V G V VG+RT I PDGLH+IPNRNI VSN SR NMRA +DI
 Sbjct: 127 QIDVGDVITAGGVGIVEEGLRTALIRGFDGLHYIPNRNIANVSNHSGNMRALVDIS 186

25 Query: 190 LFPVTLNLDQISDIVTKINEEYVSKHPAIVEGPTVPGPTINANGQFVYRINIFTQNGAQFD 249
 + + N+D+ ++ K+ ++ + I+ P V G + V RI T+ N Q+
 Sbjct: 187 ISYNDNIDEALISVMQVCDQLAQDERITIEGPDVIGVQNLGDSVDVIRI IAKTENMEQMS 246

Query: 250 IYAEFYKLYKAILLEGIDLP 270
 + K ++A+ I++P
 Sbjct: 247 VERLLRKQLKEALEAHNIETP 267

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6903> which encodes the amino acid sequence <SEQ ID 6904>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

35 INTEGRAL Likelihood = -8.49 Transmembrane 24 - 40 (15 - 45)
 INTEGRAL Likelihood = -4.83 Transmembrane 78 - 94 (73 - 99)
 INTEGRAL Likelihood = -2.07 Transmembrane 96 - 112 (95 - 113)

----- Final Results -----

40 bacterial membrane --- Certainty=0.4397(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 104/249 (41%), Positives = 151/249 (59%), Gaps = 4/249 (1%)

45 Query: 22 KKLVSLLIILMLFFAILKRVNIVLFRKTIKNSFAYSRQSBARKKTLKLVNINLYLLYPL 81
 K LV+I L+ AI KR+ + F + + + S R TL KL+ N +Y L F+
 Sbjct: 23 KVLVAVIAFLIVRAIGKRIISNSFARMAKNN----QLSSGRVVTLEKLSINAFSYTLMF 78

50 Query: 82 LIYWLISLFGIPVSSLLAGAGIAGVAILGAAGQGLSDVNVNGFTILENPFQVGRNIVTSD 141
 +L+FG+ S+L+AGAGI G+AI G GAQG +SD+V GPFILE E Q +VD V T
 Sbjct: 79 FATTLLITPGLNPSALAGAGIAGVILGAIGPGAQGLVSDIVTGVFTILEKQIDVGDVYTAGG 138

55 Query: 142 IEQSVFGVGIRITQIHDFDGLHFIIPNRNITVSNKSRNMRALQIDIPVYRINIFTQNGAQFD 201
 +G V VG+RT IRGPDGLH+IPNR+I VSN SRNMRL++I + N+ +
 Sbjct: 139 VDIGVEEGLRTALIRGFDGLHYIPNRNIANVSNHSGNMRALVDISYNDNIDEAIS 198

60 Query: 202 IIDEVQKELFNPHTQIVQKPNILGPNQNNNGQFTFRIATFENGEQFPIYHFFRYLQRA 261
 ++ +V + +I+ P+++G QN + RI TEN EQ+ + + +EA
 Sbjct: 199 VMQVCDQLAQDERITIEGPDVIGVQNLGDSVDVIRI IAKTENMEQMSVERLLRKQLKEA 258

Query: 262 LLKEGIQLP 270

-2521-

L I++P
 Sbjct: 259 LEAHNIEIP 267

An alignment of the GAS and GBS proteins is shown below.

Identities = 164/265 (61%), Positives = 215/265 (80%)

Query: 7 FIDHLNVEBVLFTFFTKLISILLIIAPVIVRQVINLFEKTVNRSFLAFSRQKVARQKTL 66
 +++ ++E + T F KL+S++L++ P I++V NYLFEXT+N+S A+SRQ AR+KTL
 Sbjct: 7 YLQGHIEINIGLTI PKKLVSLLILLFFAILKRVINLFEKTVNRSFLAFSRQKVARQKTL 66

Query: 67 AKLSHNVINNTLYFFLFYWLISILGVPISSLLAGAGIAGVAIGLGAQGFLSDVNVGFFIL 126
 +KL+HN+IN YLF L YWLIS+ G+P+SSLLAGAGIAGVAIGLGAQGFLSDVNVGFFIL
 Sbjct: 67 SKLTINILNLYLFLIYWLISLFGIPVSSLLAGAGIAGVAIGLGAQGFLSDVNVGFFIL 126

Query: 127 LKQGFVAGDIINVGTVSGTIVNVGIRTTQIHDFDGLHFIPNKNITVSNKSRNMRQI 186
 ENQF+VGD + + + G+V VGI RTTQI FDTLHFIPNR+IT+VSNKR NMRA I
 Sbjct: 127 FENQFEVGDVNTISDIESSVFGVGI RTTQIRGFDGLHFIPNRSITVSNKSRNMRALI 186

Query: 187 DIPLFVHTNLDQISDIVTKINEEYVSKHPAIVGEPTVFGPTNANGQFVYRINIFTQNGA 246
 +IPL+ NL Q+ I+ ++N++ + HP IVG+P + GP N+NGQF +RI IPT+NG
 Sbjct: 187 EIPLYSTVNLISQVTRIIDEVQKELPNHPQIVGKPNILGPNNSNGQFTFRIAIPTENG 246

Query: 247 QFDIYAEFYKLYQKALIEEGIDLEPT 271
 QF IY FY+LYQ+A+L+EGI LPT
 Sbjct: 247 QFKIYHTFYRLYQRAILKEGIDLEPT 271

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2235

- 30 A DNA sequence (GBSx2354) was identified in *S. agalactiae* <SEQ ID 6905> which encodes the amino acid sequence <SEQ ID 6906>. This protein is predicted to be RopA (tig). Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1785(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9283> which encodes amino acid sequence <SEQ ID 9284> was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6907> which encodes the amino acid sequence <SEQ ID 6908>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0776(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 303/354 (85%), Positives = 337/354 (94%)

Query: 1 MSTSPENKATNRGIIITFTISQDEIKPALDQAPNKVKIDLVNPGFRKGHMPRTVFNQKPG 60

-2522-

```

MSTSFENKATNRG+ITFTISQD+IKPALD+AFNK+KKDLN  PGFRNGHMPR  VFNQKQFGE
Sbjct: 30  MSTSFENKATNRG+ITFTISQDKIKPALDKAAPNKIKKDLNAPGFRNGHMPRPFVFNQKQFGE  89

Query: 61  EALYENALNIVLPKAYEAAVDELGLDVAQFKIDVVSMEKQDKWKLTAESVVTKEPVLKGD  120
      E LYE+ALN+VLP+AYEAAV  ELG:LDVVAQFKIDVVSMEKQ++W  L+AEVVTKEPVLKGD
Sbjct: 90  EVLVEALNIVLPKAYEAAVTELGLDVAQFKIDVVSMEKQEWTLASAEVVTKEPVLKGD  149

Query: 121  YKDLSEVDASKEVSDSEVDIAKVERERNNAELTVKDGAAQGEVTVIDFVGSVDGVSEFD  180
      YK+L  VEVDASKEVSD+VDAK+ERRR  NLAE  +KDGEAAQGEVTVIDFVGSVDGVSEFD
Sbjct: 150  YKDLSEVDASKEVSDSEVDIAKERRERQNLAEIITKDGAAQGEVTVIDFVGSVDGVSEFD  209

Query: 181  GKGNDFSELELGGGQFIFGFEQLVGSKAGQTVDVNVTFPEDEYQARDLAGKDAKVFVTIH  240
      GKGNDFSELELGGGQFIFGFE+QLVG+KAG  V+VNVTFPE  YQARDLAGK AKF+TTIH
Sbjct: 210  GKGNDFSELELGGGQFIFGFEQLVGAAGDEVEVNVTFPESEYQARDLAGKAAKFMVTIH  269

Query: 241  EVKTEKVPALDDELAKDIDDSVETLDELNAKYRKELESARKEIAPDDAVEGAALAVANA  300
      EVKTEKVP  LDELAKDID+V+TL++LK  KYRKELE+A+S  A+DDAVEGAALAVANA
Sbjct: 270  EVKTEKVPALDDELAKDIDSDVITDELVKYRKELEAAGETAYDDAVEGAALAVANA  329

Query: 301  EIVLPEEMVHDEVHRAMNFMGNWQKQGISPENITFQLTGTTEEDLHNYQADA  354
      EIV+LPEEM+H+EV+R++NEFMGNWQKQGISPENITFQLTGTIT+EDLH  QY  A+A
Sbjct: 330  EIVLPEEMTHDEVNRVNEFMGNWQKQGISPENITFQLTGTITCEDLHNYQADA  383

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2236

A DNA sequence (GBSx2355) was identified in *S.galactiae* <SEQ ID 6909> which encodes the amino acid sequence <SEQ ID 6910>. This protein is predicted to be galactose-6-phosphate isomerase laca subunit (rpiB). Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3491(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA25177 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
Identities = 92/141 (65%), Positives = 115/141 (81%)

Query: 1  MTILIGADAHGVLEKEVIRQHLTSLGKRIIDLITDSKDFVDNTIAIVAKVNCKEDNLGIM  60
      M I++GAD  G  LK+V++  L  G  R+ID+T  +DFVD  TLA+  ++VN+  E  NLGI+
Sbjct: 1  MAIVVGADLKGTRLEKDVVKNFLVEEGFEVIDVTKDGQDFVDVITAVASEVNKDEQNGLIV  60

Query: 61  VDAYGVGPFPMVATKIKGMIAAEVSDERSAYMTRAHNNARMITLGEIVKGVAKHIVBGF  120
      +DAYG  GPFPMVATK+KGM+AAEVSDERSAYMTR  HNNARMIT+G+EIVG  +AK+I  + F
Sbjct: 61  IDAYGAGPFPMVATKIKGMVAAEVSDERSAYMTRGHNNARMITVQAEIVGDELAKNIAP  120

Query: 121  VDGTYDAGRHOQIRVDMNLNKC  141
      V+G  YD  GRHO+RVDMNLNKC
Sbjct: 121  VNGKYDGRHOQIRVDMNLNKC  141

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6911> which encodes the amino acid sequence <SEQ ID 6912>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

-2523-

bacterial cytoplasm --- Certainty=0.3224 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 5 An alignment of the GAS and GBS proteins is shown below.

Identities = 101/140 (72%), Positives = 117/140 (83%)

Query: 1 MTIIIGADAHGVELKRVIRHLSLGKEIIDLTDTSKDFVDNTLAIVAKVQKEDNLGIM 60
 M I I+GADAHG LKR+I+ L G +I I+T D + D F+DNTI A+ VN+ E L G I M
 10 Sbjct: 1 M A I L G A D A H G N A L K E L I K S P L Q R G Y I I D V T D I N S D F I D N T L A V A K A N E A B G R L G I M 60

Query: 61 VDAYGVGPFMVATKVKGMIAAEVSDERSAYMTRHNNRMITLGSEIVGPGVAKHIVEGF 120
 V D A V G G P F M V A T K V K G M I A A E V S D E R S A Y M T R H N N R M I T L G S E I V G P G V A K H I V E G F
 15 Sbjct: 61 V D A Y G A G P F M V A T K L G M V A E V S D E R S A Y M T R G H N N R M I T I G A E I V G P E L A K N I V E G F 120

Query: 121 VDGTYDAGRHOIRVEMLNHN 140
 V G Y D G R H Q I R V E M L N H N
 Sbjct: 121 V T G P Y D G G R H Q I R V E M L N H N 140

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2237

A DNA sequence (GBSx2356) was identified in *S.agalactiae* <SEQ ID 6913> which encodes the amino acid sequence <SEQ ID 6914>. This protein is predicted to be galactose-6-phosphate isomerase lacc subunit (rpiB). Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2511 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 10189> which encodes amino acid sequence <SEQ ID 10190> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAA25178 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
 Identities = 138/171 (80%), Positives = 157/171 (91%)

40 Query: 10 MKIAVCGDHIVTDKIAVVDYLTGKGYEVIDOOTYDNIRKTHYPYIGKVKGEAVASQADL 69
 M+IA+GCDHIVT K+AV ++LK+KGVEV+D GTYD++RTHYPIYGVKVKGEAV SQ+ADL
 Sbjct: 1 M R I A I G C D H I V I D V K M A V S E F L K S K G Y E V L D P G T Y D H V R K T H Y P I Y G K V K G E A V S Q A D L 60

45 Query: 70 GVCICGTGVGINAVNKVPGVRSALVRDLTSALYAKEELNANVIGFGGKITGGILMTDII 129
 G V C I C G T G V G I N A V N K V P G V R S A L V R D L T S A L Y A K E E L N A N V I G F G G K I T G G I L M T D I I
 Sbjct: 61 G V C I C G T G V G I N A V N K V P G V R S A L V R D M T S A L Y A K E E L N A N V I G F G G M I T G G I L M D I I 120

Query: 130 EAFIARAKYKPTKENKVLIRKIAEVETHNAHQEENDPFTEPLDKMNRGEYHD 180
 E A F I A R A K Y K P T K E N K V L I R K I A E V E T H N A H Q + + P P T E P L + K W + R G E Y H D
 50 Sbjct: 121 E A P I E A Y K I P T E N K K L I A K I E H V E T H N A H Q A D E R F P T E P L E K M D R G E Y I D 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6915> which encodes the amino acid sequence <SEQ ID 6916>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

-2524-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3048(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 136/171 (79%), Positives = 160/171 (93%)

Query: 10 MKIAVGCDDHIVTYDKIAVVDYLKTKGYSEVIDCGTYDNIRTHYPIYGGKVGAEVAGSKADL 69
 MKIA+GCDHIVT +K+AV D+LK+KGY+VIDCGTYD+ RTHYPI+GKKVGAEV +G+ADL
 Sbjct: 2 MKIAIGCDHIVTNERGAVSDFLKSGYVIDCGTYDHTRTHYPIFGKKVGAEVNVQADL 61

Query: 70 GVCICTGTVGINNAVNVKPGIRSALVRDLTSIAYAKREINANVIGFGKTIIGLLMTDII 129
 GVCICTGTVGINNAVNVKPGIRSALVRD+T+A+YAKEINANVIGFGKTIIGLLMTDII 129
 Sbjct: 62 GVCICTGTVGINNAVNVKPGIRSALVRD+T+YAKEINANVIGFGKTIIGLLMTDII 121

Query: 130 EAFIRAKYKPTKENKVLIEKIAEVSTHNAQRENDFFTEFLDKVNRGEYHD 180
 +AFI+A+YK T+ENK LI KIA +S+H+A+QE+ DFFTEFL+KW+RGEYHD
 Sbjct: 122 DAFIKABYKETEENKKLIAKIAHLESHRANQRENDPFFTEFLDKVNRGEYHD 172

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2238

A DNA sequence (GBSx2357) was identified in *S. agalactiae* <SEQ ID 6917> which encodes the amino acid sequence <SEQ ID 6918>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10187> which encodes amino acid sequence <SEQ ID 10188> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAA25179 GB:M60447 tagatose 6-P kinase [Lactococcus lactis]
 Identities = 192/310 (61%), Positives = 236/310 (75%)

Query: 11 MILTVTLNFSIDISYCLENFNMDTVNKVIDVSKTPGGKGLNVIRVLSQLDGNVAVTGLLG 70
 MILTVTLNFS+DISY LE +DTVNK VDSKT GSKGLNVIRVL +GD V ATG LG
 Sbjct: 1 MILTVTLNFSVIDISYPLETLKIDTVNKRVIDVSKTAGGKGLNVIRVLGVESGDKVATGFLG 60

Query: 71 GDFGDFIRSGLDALRIHQPLSIGGETRHCIAVLH83QTEILEKGPHITKDERDAFLSH 130
 G G+FI S L+ + F I G TR+CIA+LH8G QTEILE+GP I+ +EA+ FL+H
 Sbjct: 61 GKICEFIESLEQSPVSPAPYKISGNTRNCIALH8GNQTEILEQGPTISHERAGFLDH 120

Query: 131 LKLFIDAA+IITVSGSLFKGLPSDYARLISLANHFNKKVVLDCSGEALRSVLKSSAKPT 190
 + + +T+SGSLP GLP+DY +LI LA+ VVLDGSG L +VLKSSAKPT
 Sbjct: 121 YSNLIKQSEVVTISGSLPSGLPNDYTKELIQASDEGVAVVLDGSGAFLSVLKSSAKPT 180

Query: 191 VIKNLEELTQLIGKPIYSYSLDELKSTLQQLFRGIDWVTVSLGARGAPAKHNNHYQVT 250
 IKFN EEL+QL+GK ++ ++ELK L++ IF GI+W++VELG GAFARHG+ +Y+V
 Sbjct: 181 AIKPNLEELSOLLGKEVTKDIEELKQVLKLESLGIEWIVSLGRNGAFARHGDFVYK 240

Query: 251 IPKIEVINPVGSGEATVAGIASALEHQLDTHLLKRVNLGMLNAQETLTHGINLTYTQ 310
 IP I V+NPVGSDE+TVAGIASAL + D +LLK A LGMLNAQET+TGH+N+T Y+
 Sbjct: 241 IPDIPVNVFVGSDETVAGIASALNSKSDADLLKHAMITGLMLNAQETWTHVNMNITYET 300

-2525-

Query: 311 LISQIQVKRV 320
 L SQI VKRV
 Sbjct: 301 LNSQIQVKRV 310

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6919> which encodes the amino acid sequence <SEQ ID 6920>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1178 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 184/310 (59%), Positives = 232/310 (74%), Gaps = 1/310 (0%)

- Query: 11 MILITVLNPSIDISYCLNFNMOTVNRVTVDSKTFGGKGLNVTRVLSQLGDNVVATGILG 70
 +ILITVLNLP+ID+SY L+ DIVNRV DV+KTGGKGLNV+RVL++ G+ V NTG +G
 20 Sbjct: 1 VILITVLNPAIDVSYFLDELKCDTVNRVVDVTKTFGGKGLNVSRVLEPGETVKATGCGV 60
- Query: 71 GDFGDFIRSGLDALIRHQFLSIGGETRHCIAVLHBGQQTILEKGFPHITKDEADAFINR 130
 G+ GDFI + L I +F I G+TR CIA+LHBG QTEILENGP ++ DE D F +H
 25 Sbjct: 61 GESGDFIINHLPD-SILSRFYKISGDTKCTCAILEBGNQTEILENGFMLSVDEIDGFTH 119
- Query: 131 LKLIIFDRAITITVSGSLPKGLPSDYARLISLANHFNKKVVLDCSGEALASVLKSSANPT 190
 K + + ++T+SGSLP G+F DYY +DI +AN KK VLDCSG AL +VLK +RPT
 30 Sbjct: 120 FKYLLNDVNDVVLGSLPAGMFDYYQKLIKIANLANGKKTVLDCSGNALEAVLKGDSKPT 179
- Query: 193 VIKFNLBELTQLIGKPISTYSLDELKSTLQQDLFRGIDWIVISLGARGAPAKHGHNYQVT 250
 VIKFNLBEL+QL+GK ++ D LK LQ +LF GI+H+IVSLGA G PAKH + +Y V
 35 Sbjct: 180 VIKFNLBELSQLSKEMTKDFALKEVLQDELFDGIEWIIIVSLGADGVFAKHKDTFNVND 239
- Query: 251 IPKIEVINFGSGDATVAGIASALEHQDLDTNLLKRAVFLGMLNAQETLTGHINLTYQSR 310
 IPKI++++ VSGSD+TVAGIAS L + DD LL +ANVLGMLNAQE TGH+H+ Y +
 40 Sbjct: 240 IPKIKIVSAVSGSDTVAGIASGLANDEDDRALLTYANVLGMLNAQETTGHRVNMANYDK 299
- Query: 311 LISQIQVKRV 320
 L I+VKRV
 40 Sbjct: 300 LYSQIRVKRV 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2239

- 45 A DNA sequence (GBSx2358) was identified in *S.agalactiae* <SEQ ID 6921> which encodes the amino acid sequence <SEQ ID 6922>. This protein is predicted to be tagatose 1,6-diphosphate aldolase. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

- 50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0369 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 55 The protein has homology with the following sequences in the GENPEPT database.

>GP:AA25180 GB:M60447 tagatose 1,6-diP aldolase [Lactococcus
 lactis]

-2526-

Identities = 253/325 (77%), Positives = 295/325 (89%)

Query: 1 MGLTEQKQKIMBQLSDKNGIISALAFDQRGALKRIMAKYQSEPTVQIEALKVLVAEEL 60
 5 Sbjct: 1 MVLTEQKRSLEKLSDKNGFISALAFDQRGALKRIMAAQYQTEPTVQIMBELKVLVADEL 60

Query: 61 TPYASSMLLDPEYGLPATKVLDDWAGLLIAYEKTGYDTSSTKRLPDCLDIWSAKRIKESG 120
 10 Sbjct: 61 TKYASSMLLDPEYGLPATKLDKENGALLIAPFKTYDTSSTKRLPDCLDIWSAKRIKESG 120

Query: 121 ADAVKFLLYYDVDSSEVNEKEAYIERIGSECVASDIPFFLEILSYDEKITSGGIYA 180
 Sbjct: 121 ADAVKFLLYYDVDSSEDE-N++K+AYIER+GSCVASDIPFFLEIL+YDE+I+D+ +SYA 180

Query: 181 KIKPRKVIEMKVFSPNPRFNIDVLKVEVPVNMIPVSGFAQGEYAKATAAAYFREQDQA 240
 15 Sbjct: 181 KVKPRKVIEMKVFSPNPRFNIDVLKVEVPVNMKVTSFGADGEVYKAEADFFKAQESA 240

Query: 241 TLLPYIFLSAGVPAQLFQSTLVPAKEGAKFNGVLCGRATWAGSVKEVKEGAGARQWL 300
 20 Sbjct: 241 TLNPLYLSAGVSAKLFQSTLVPAHDSGAKFNGVLCGRATWAGSVKEVKEGAGAREWL 300

Query: 301 RTIGFQNIDELANKILQNTATSWKER 325
 25 Sbjct: 301 RTTGPNIDELANKVLVTSAPWTK 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6923> which encodes the amino acid sequence <SEQ ID 6924>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0600 (Affirmative) < succ>
 35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 230/323 (71%), Positives = 276/323 (85%), Gaps = 1/323 (0%)

40 Query: 3 LTEQKQKIMBQLSDKNGIISALAFDQRGALKRIMAKYQSEPTVQIEALKVLVAELTP 62
 Sbjct: 5 LTKRKRSLEKLSDKNGFISALAFDQRGALKRIMAAQYQTEPTVQIMBELKVLVADELTP 63

45 Query: 63 YASSMLLDPEYGLPATKVLDDWAGLLIAYEKTGYDTSSTKRLPDCLDIWSAKRIKESGAD 122
 +ASS+LLDPEYGLPA++V +AGLLIAYEKTGYD +T RLDCD+D+WSAKRIK GA+
 Sbjct: 64 FASSILLDPEYGLPASVRSEAGLLIAYEKTGYDMMTTSRLPDCLDIWSAKRIKESGAE 123

Query: 123 AVKFLLYYDVDSSEVNEKEAYIERIGSECVASDIPFFLEILSYDEKITSGGIYAKI 182
 AVKFLLYYD+D +VNE+K+AYIERIGSECVASDIPFFLEIL+YDEKI D++ E+AK+
 50 Sbjct: 124 AVKFLLYYDIDSDQVNEBKAYIERIGSECVASDIPFFLEILSYDEKIDNASPEFAKV 183

Query: 183 KPRKVIEMKVFSPNPRFNIDVLKVEVPVNMIPVSGFAQGEYAKATAAAYFREQDQNTL 242
 K KV EAMKVFSP RF +DVLKVEVPVNMIPVSGFAQGEYAKATAAAYFREQDQNTL 242
 Sbjct: 184 KAKKVIEMKVFSPKPRFQDVLKVEVPVNMIPVSGFADGEVLFTEKEMQAQFREQDQASTD 243

55 Query: 243 LPYIFLSAGVPAQLFQSTLVPAKEGAKFNGVLCGRATWAGSVKEVKEGAGARQWLRT 302
 LPYI+LSAGV A+LQ+TLVFA E+GAKFNGVLCGRATWAGSVK Y+E+G AR+NLRT
 Sbjct: 244 LPTYI+LSAGVSAKLFQSTLVFAAESGAKFNGVLCGRATWAGSVKYYTEGPAAREWLRT 303

60 Query: 303 IGPFQNIDELANKILQNTATSWKER 325
 GP+NIDELANK+L KTA+ W E+
 Sbjct: 304 EGPKNIDELANKVLDTASPWTEK 326

-2527-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2240

A DNA sequence (GBSx2359) was identified in *S.agalactiae* <SEQ ID 6925> which encodes the amino acid sequence <SEQ ID 6926>. This protein is predicted to be lacx protein, chromosomal. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0643 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10185> which encodes amino acid sequence <SEQ ID 10186> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25184 GB:M60447 ORF [Lactococcus lactis]
Identities = 173/298 (58%), Positives = 219/298 (73%)

Query: 24  MATTIQHSHLQVTLKALGATMTSITDSQGVYLAQGDATYWGQQAPILFPIQGSVRNDCV 83
      M I ++N L V K LG +TSI D G+EYIMQ D YW QQAPILFP+IGS+RND
Sbjct: 1  MTIELKNGEYLTVQFKTLGSGQLTSIKDKDGLSYLAQADPEYWNQAPILFPIQGSRLNDGA 60

Query: 84  IYRPAQAPHFIGIIPRHGFVRHKTFDYDYSISVSRFTIKSSKEMLINYPYRFSLEIITYT 143
      IYRP + P FIG+I RHGFVR + F + +++SV F+IK + EML NY Y+P L + YT
Sbjct: 61  IYRQBRPFPTGLIRRHGFVRKEFTLEEVNENSVTFSLKPNARMLDNLYQFELRVVYT 120

Query: 144 LRNKSHIAITYIVKNLSSEKNMPPYAIGAHPGFCNPLFEKEVPDSYYLEFQFETCTIPESF 203
      L NSI + V NLE+EK MPY IGAHP FNCPL E R + DY LEF + E+C+IP+SF
Sbjct: 121 LNKGSIRTEPQVTNLSTETKMPYFIGAHPFNCPLVBGRKYEDYSLEFSEVESCSIPKSF 180

Query: 204 PGTGLLDLQARHPFLENQKQLSLMHAFKEDAITLDQLSKTYYLKSRLNHAQIQLQDFDD 263
      P+TGLLDLQ R PFLENQK L L+++LF DATILD+L+S++V L+SR KG+++DFDD
Sbjct: 181 PETGLLDLQDRTPFLENQKSLDLYSLFSDAITLDLRLKSRSVTLRSRFGSGRLVDFFD 240

Query: 264 FENLILVTSNNGGPFIALEFPWSLSTSIIESDILEDKQNIIVLNPNKQSKQHSIRITIL 321
      F NLILN++ N PF+ALEFWS LST+EE +ILEDK + ++ P + + S ITIL
Sbjct: 241 FENLILVSTTNKSPFIALEFPWSLSTSLSEGNILEDKQNVKLPVLPDTSKESYDITIL 298
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2241

A DNA sequence (GBSx2361) was identified in *S.agalactiae* <SEQ ID 6927> which encodes the amino acid sequence <SEQ ID 6928>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3272 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
```

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10183> which encodes amino acid sequence <SEQ ID 10184> was also identified.

- 5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAAS1350 GB:X72832 leucine rich protein [Streptococcus
equisimilis]
Identities = 101/278 (36%), Positives = 160/278 (57%), Gaps = 1/278 (0%)

10 Query: 10 MDPKELFPEVKITQKQSEDIYIVQDGHVHFPKSSLTKHLYLLQMPFSLDASSVD 69
M+ K+ FFE+ +++ V++ +HPPKS L+++E LL++ +
Sbjct: 1 MELKDYFPEMQVGHPIGLDKEWVSVKRSDQVHFPKSCSLSEKRLLEVLGLQYVELQ-P 59

15 Query: 70 SQNFWRYLVBGRGRLPQSHSAVQPIFIEHQPTLSEELADPLSLVINVETIMTINQTS 129
+FW RYL++ +G PQ QFI++ HQ L +L +L ++ +E I+ I+ TQ+
Sbjct: 60 LGSFWQRYLLDHQGNPPQLFETSQFIYLNHQVLPADLVLLQMGVLEVLPISTTQT 119

Query: 130 VMILNQCNFFNATELLTDILPTIENDFNTRLCYPCGNSWHLQAVDWKELYEEYKLPFL 189
+ Q L +LPT+E+DF L + GN+W + A +E +EE +L T
20 Sbjct: 120 AFLCRQATSIKVLRSLEGLLPTLSDPGLALTMFVGNAMYQVAGTLRECFREKQLLTA 179

Query: 190 FLSHKABQHYCRFPKPMALNALNQSPMPSIKAKCLQHILDTSDTSIAIKALNQSQNLAK 249
+L K+ F ++ LW++ + P++ + Q + SD + ++ ALN E GNL +
25 Sbjct: 180 YLKQKSGKLLTFPQVLMNSILSHQSPFALTRQPHQFLNQSMDADVHLANSEHGNLQ 239

Query: 250 TAKALFTHNSLYQKLDKFTQSSGILNKILDDLAAYNL 287
TA+ L+THNSLYQKLDKF Q SGL+LK LDDLA+AYL
Sbjct: 240 TAQRLYTHNSLYQKLDKFAQSSGLHLKQLDDLAFAYL 277

- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6929> which encodes the amino acid sequence <SEQ ID 6930>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4332 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 40 An alignment of the GAS and GBS proteins is shown below.

Identities = 106/287 (36%), Positives = 169/287 (57%), Gaps = 4/287 (1%)

Query: 3 KTVVSD-AMDPKELFPEVKITQKQSEDIYIVQDGHVHFPKSSLTKHLYLLQMP-TP 60
KTV++ AM+ K+ FFE+ +D++ +++ +HPPKS L+++E LL++
45 Sbjct: 7 KTVKGMGMELKDYFPEMQVGHPIGLDKNWSIKRSDQVHFPKSCSLSEKRLLEVLGLG 66

Query: 61 SLEDASSVDGQSNFWRYLVBGRGRLPQSHSAVQPIFIEHQPTLSEELADPLSLVINVET 120
E +S FW RYL++ +G PQ + QFI++ HQ L +L +L ++ +E
50 Sbjct: 67 QCEVIQPLGS--FWQRYLLDHQGNPPQLFETSQFIYLNHQVALPDVLVLLQMGVLE 124

Query: 121 IMTINQTSVMILNQCNFFNATELLTDILPTIENDFNTRLCYPCGNSWHLQAVDWKELY 180
I+ I+ TQ+ + Q L D+LPT+E+DF L + GN+W + A +E +
Sbjct: 125 ILPISATQTATLCRQAISIKVLRLLEDLLPTLSDPGLALTMFVGNAMYQVAGTLRECF 184

55 Query: 181 EEEYKLPFLFLSHKABQHYCRFPKPMALNALNQSPMPSIKAKCLQHILDTSDTSIAIKAL 240
EEE +L T +L + + + F + LW+L + ++ + Q + SD + ++ AL
Sbjct: 185 EEEQQLLTAYLRQSGKLLTFPGLMLNSLSHFTPLALTRQPHQFLNQSMDADVHLA 244

Query: 241 WQBGNLAKTAKALFTHNSLYQKLDKFTQSSGILNKILDDLAAYNL 287
W E GNL +TA+ L+THNSLYQKLDKF Q SGL+LK LDDLA+AYL
60 Sbjct: 245 WSEHGNLQVTAQRLYTHNSLYQKLDKFAQSSGLHLKQLDDLAFAYL 291

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2242

A DNA sequence (GBSx2362) was identified in *S. agalactiae* <SEQ ID 6931> which encodes the amino acid sequence <SEQ ID 6932>. This protein is predicted to be multiple sugar-binding transport ATP-binding protein msmk (malK). Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4392 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26938 GB:M77351 ATP-binding protein [Streptococcus mutans]
Identities = 320/377 (84%), Positives = 359/377 (94%)

Query: 1 MVELNLNHIYKYPASHSYSEVDFDLIDKKEFIVFVGPGCGCKSTTLRMAGLEDISEG 60
MVELNLNHIYKYP++SHYSVEVDFDLIDK+KEFIVFVGPGCGCKSTTLRM+AGLEDI++G
Sbjct: 1 MVELNLNHIYKYPNSSHSYSEVDFDLIDKKEFIVFVGPGCGCKSTTLRMVAGLEDITKG 60

Query: 61 ELKIDGEEVNDKSPKDRDIAMVFQNYALYPIMTVYDNNMAGFLKLRKFSKEIDKRVREA 120
ELKIDGEEVNDK+PKDRDIAMVFQNYALYPIM+VIDNNMAGFLKLR ++SK+ IDKRV+EA
Sbjct: 61 ELKIDGEEVNDKAPKDRDIAMVFQNYALYPIMSVYDNNMAGFLKLRHSKEAIDKRVREA 120

Query: 121 ANIGLTFLEKRPADLSGSGQRQVAMGRAIVRDAKVFLMDEPLSNLDKLRVSMRAEIAK 180
+GLTFLEKRPADLSGSGQRQVAMGRAIVRDAKVFLMDEPLSNLDKLRVSMRAEIAK
Sbjct: 121 QILGLTFLEKRPADLSGSGQRQVAMGRAIVRDAKVFLMDEPLSNLDKLRVSMRAEIAK 180

Query: 181 IHQRIGSTTIYVTHDQTEAMTLADRIVMSATNPDSGDTIGKIEQVGSQELYNLPANK 240
IH+RIQ+TTIYVTHDQTEAMTLADRIVMS+TN DG GTIG++EQVG+QELYN PANK
Sbjct: 181 IHRRIQATTIYVTHDQTEAMTLADRIVMSSTINEDSGDTIGRVEQVGTPQELYNLPANK 240

Query: 241 FVAGFIGSPSMNFFVKVKNMGLISEDGLRIAIPGQKLLSRGYKGLIFGIRPEDI 300
FVAGFIGSP+MNFF V +++G ++S+DGL IA+ EQG K+LES+G+K K LI FGIRPEDI
Sbjct: 241 FVAGFIGSPAMNFFVDITKDHILVSKDGLTIAVTEGQLMLSESGKGFNGLIFGIRPEDI 300

Query: 301 SSNLLVQDTYPNANVEASVLVSELLGSETMLYVKLGQTEFAARVDARPHNPGEKVNLTF 360
SS+LLVQ+TYP+A V+ASV+VSELLGSETMLY+KLGQTEFA+RV+ARDPH PGEKV+LTF
Sbjct: 301 SSSLVQDTYPDATVASVSVSELLGSETMLYKLGQTEFAARVDARDPHEPGEKVSLTF 360

Query: 361 NVAKGHFFDADETAIR 377

NVAKGHFFDA+TE AIR

Sbjct: 361 NVAKGHFFDAETEAIR 377

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6933> which encodes the amino acid sequence <SEQ ID 6934>. Analysis of this protein sequence reveals the following:

Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4642 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 332/377 (88%), Positives = 359/377 (95%)

-2530-

- Query: 1 MVELNLNHIYKYPASHSYSDVDPLDIDKKEFIVFVPGSGCGKSTTLRMIAGLDEISEG 60
 MVELNLNHIYKYP+ +HY+VEDFDLIDKKEFIVFVPGSGCGKSTTLRMIAGLDEISEG
 Sbjct: 1 MVELNLNHIYKYPWTHYAVSDFDLIDKKEFIVFVPGSGCGKSTTLRMIAGLDEISEG 60
- Query: 61 ELKIDGEVNDKSPKDRDIAMVQNYALYPHMYVYNMAGLKLKRPKSQEKDRVREAA 120
 ELKI GEVNDKSPKDRDIAMVQNYALYPHMYVYNMAGLKLKRPK+ K+ID+RV+RAA
 Sbjct: 61 ELKIDGEVNDKSPKDRDIAMVQNYALYPHMYVYNMAGLKLKRPKDDIDRRVREAA 120
- Query: 121 ANIGLTFELERKPADLSGGQRQVRVAMGRAIVROAKVFMDEPLSNLDAKLRVSMRAEIAK 180
 +GLTSEFLERKPADLSGGQRQVRVAMGRAIVROAKVFMDEPLSNLDAKLRVSMRAEIAK
 Sbjct: 121 QILGLTFELERKPADLSGGQRQVRVAMGRAIVROAKVFMDEPLSNLDAKLRVSMRAEIAK 180
- Query: 181 IHRIGSTTIYVTHDQTEAMTLADRVIMSNTRNEDSGTIGKIEQVGSQPELYNLNPAK 240
 IR+RIGSTTIYVTHDQTEAMTLADRVIMSNTRNEDSGTIGKIEQVGSQPELYNLNPAK
 Sbjct: 181 IHRIGSTTIYVTHDQTEAMTLADRVIMSNTRNEDSGTIGKIEQVGSQPELYNLNPAK 240
- Query: 241 FVAGFIGSPSMNFFKVKVENGMIISEDGLRIAPBQEKLLSESGYKGLKIFGIRPDI 300
 FVAGFIGSP+MNFV+V+V+G I+SEDGL IALPESG K+LE+ GYNG+++ FQIRPDI
 Sbjct: 241 FVAGFIGSPAMNFFVEVDKGRIVSESDGLIAPBEGAGMLEAANGYKGVKVFQIRPDI 300
- Query: 301 SSNLLVQDTYPNANVAEVLVSELLGSETMLYKLGQTEFASRVARDPHNPKGNLTF 360
 SS +V DTYP+A V AEVLVSELLGSETMLYKLGQTEFASRV+ARDPH+PGB+V+LTF
 Sbjct: 301 SSRQIVHDTYPSATVTAEVLVSELLGSETMLYKLGQTEFASRVARDPHSPGBQVSLTF 360
- Query: 361 NVAKGHFFDADTEQAIR 377
 NVAKGHFFD DTEQAIR
 Sbjct: 361 NVAKGHFFDRITBQAIR 377
- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2243

A DNA sequence (GBSx2363) was identified in *S. agalactiae* <SEQ ID 6935> which encodes the amino acid sequence <SEQ ID 6936>. This protein is predicted to be glucan 1,6- α -glucosidase (dexB) (trcC).

- Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

- Final Results -----
 bacterial cytoplasm --- Certainty=0.2525 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- >GP:CAAS1348 GB:X72832 glucan 1,6- α -glucosidase [Streptococcus equisimilis]
 Identities = 413/535 (77%), Positives = 476/535 (88%), Gaps = 1/535 (0%)
- Query: 1 MKKHWHKATIIYIYPRSPMSDGLGVGDIKITSKIDVLEKLGITATLWLSPPVQSPMD 60
 M+K WHKATIIYIYPRSF D+ G+G+GD+KGITS+LDYL+KLGITATLWLSPPVQSPMD
 Sbjct: 1 MQQWMIKATIIYIYPRSPFDVSGNIGDLKITSQIDYLQKLGITATLWLSPPVQSPMD 60
- Query: 61 NGYDISDYQALADIPGDNDMDQLQANQRGIKILINDLVNHTSDEHAMFVEARENPN 120
 NGYDISDY+ALA++FG+M+DMD LL AN+RGIKILINDLVNHTSDEHAMFVEARENPN
 Sbjct: 61 NGYDISDYEAIAEVFGNDMDNDLLAANERGIKILINDLVNHTSDEHAMFVEARENPN 120
- Query: 121 PERDFYTRDEPNLDSIFSGSAWEYDKVSGQYIHLHPSKRPDLNWRNEALSHKIYDM 180
 PERD+YTRDEPN+L SIFSGSAWE D+ SGQYIHLHPSK+QPDLNWEN+R KIYDM
 Sbjct: 121 PERDYIYTRDEPNLMSIFSGSAWELEASQYIHLHPSKRPDLNWRNEALSHKIYDM 180
- Query: 181 NFWIDKIGGFRMDVIDLIGKIPDKITGNGPKLHDYKEMNRASFGKHDLTIVGETVGA 240

-2531-

NWFI KGIGGFRMDVIDLIGKIPD ITGNGP+LHDYLKEMN+A+PG HD++TWGETWGA
 Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKIPDSEITITGNGPLHDYLKEMNQATPGMDVMTVGETWGA 240

5 Query: 241 TPDIARQYSNPINBELSMVQFQFHVGLQHKPDAPKWDYSOGLDVPALADIFTKWQTGLEL 300
 TP+IA+QYS P+N+ELSMVQFQFHVGLQHKPDAPKWDY++ LDVPALK IP+KWQT+L+L
 Sbjct: 241 TPDIARQYSRPNKELSMVQFQFHVGLQHKPDAPKWDYAEGLDVPALKITPSKWQTLEKL 300

Query: 301 QGWNLSFWNNHDLPRVLSTWGNDSNRRQSAKALAILLHMRGTPIYIQGEIGMTNYP 360
 G+WNLSFWNNHDLPRVLSTWGNDS R++SAKALAILLHMRGTPIYIQGEIGMTNYP
 10 Sbjct: 301 GEGWNLSFWNNHDLPRVLSTWGNDSITYREKSAKALAILLHMRGTPIYIQGEIGMTNYP 360

Query: 361 FECLADVDDIESLNAYAKAMNGVSEATILDSIRKVRGNARTFMQNSQHQAGFTGK-T 419
 F+L+VDDIESLNAYAKAM+NGV A++SIRKVRGNARTFMQNS++ AGF++
 Sbjct: 361 FKDLTEVDDIESLNAYAKAMNGVPAARVNSSIRKVRGNARTFMQNSKDTHAGFSEAKE 420

15 Query: 420 FWLAVNPNYQSEINVEAALNDTESIFTYQKLVLRKEHDLVDADPKLEETADKVPAYVR 479
 WL VNPNYQSEIN AL + +SIFTYQ+L+ALRK+ DMLV+AD+ LL TADKVPAY R
 Sbjct: 421 TWLAVNPNYQSEINADALANQDSIFTYQQLIALRKDQMLVADTHLLPTADKVPAYQR 480

20 Query: 480 QTDKERYILVANLSDQNSPEFFPEAVKETIISNTEVQEVLSNLTKLWDAFCIEL 534
 Q +E Y+IV N+SDQ Q F A R +I+RT+V +VL + L+PWDAFC+LL
 Sbjct: 481 QGGEETVIYVNVSDQGVPAKDLAGEVVIITNTVDKVLSTKHLQWDAFCVKL 535

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6937> which encodes the amino acid
 25 sequence <SEQ ID 6938>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2793(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 418/535 (78%), Positives = 474/535 (88%), Gaps = 1/535 (0%)

Query: 1 MKGEWHKEATIIQIYPRSFMDSDGSDGVDIKITSLDYLEKLGITAIWLSPVYQSPMD 60
 M EWHKEATIIQIYPRSP D+ G+G+GD+KGITS+LDYL+KLGITAIWLSPVYQSPMD
 40 Sbjct: 1 MNEWHKEATIIQIYPRSFKDTSGNGIGDLKGITSQLDYIQLKLGITAIWLSPVYQSPMD 60

Query: 61 NGYDISDYQAIADIFGDMNDMDQLQEAQNGRIKIMDLVNHSTDEHAWFVARENPNS 120
 NGYDISDY+AIAD+FGDM DMD+LL AN+RGIKIMDLVNHSTDEHAWFVARENPNS
 45 Sbjct: 61 NGYDISDYEAIDVFGDMADNDELLAANERGRIKIMDLVNHSTDEHAWFVARENPNS 120

Query: 121 PERDFTIWRDEPNDLTISIPSGSAWEYDKVSOQYILHLFSLKRPDLIWLWNSALRHKIYDM 180
 PERD+TIWRDEPN+L SIFSGSAWE D+ SOQYILHLFSLK+QPLIWIEN LR KIYDM
 50 Sbjct: 121 PERDFTIWRDEPNLWISIPSGSAWELDRASQYILHLFSLKRPDLIWLWNAQLRQKIYDM 180

Query: 181 NFWIDKIGGFRMDVIDLIGKIPDKGITGNGPKLHDYLKEMNRA+PG HD++TWGETWGA 240
 NFWI KGIGGFRMDVIDLIGKIPD ITGNGP+LHDYLKEMN+A+PG HD++TWGETWGA
 55 Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKIPDLEITITGNGPLHDYLKEMNQATPGMDVMTVGETWGA 240

Query: 241 TPDIARQYSNPINBELSMVQFQFHVGLQHKPDAPKWDYSOGLDVPALADIFTKWQTGLEL 300
 TP+IA+QYS P+N+ELSMVQFQFHVGLQHKPDAPKWDY+ LDVPALK IP+KWQT+L+L
 60 Sbjct: 241 TPDIARQYSRPNKELSMVQFQFHVGLQHKPDAPKWDYAEGLDVPALAKITPSKWQTLEKL 300

Query: 301 QGWNLSFWNNHDLPRVLSTWGNDSNRRQSAKALAILLHMRGTPIYIQGEIGMTNYP 360
 G+WNLSFWNNHDLPRVLSTWGNDS R++SAKALAILLHMRGTPIYIQGEIGMTNYP
 65 Sbjct: 301 GEGWNLSFWNNHDLPRVLSTWGNDSITYREKSAKALAILLHMRGTPIYIQGEIGMTNYP 360

Query: 361 FECLADVDDIESLNAYAKAMNGVSEATILDSIRKVRGNARTFMQNSQHQAGFTGK-T 419
 F+L+V+DIESLNAYAKAM NGVS A ++SIRKVRGNARTFMQNS++ AGF++
 Sbjct: 361 FKDLTEVNDIESLNAYAKAMNGVSAARVMDSSIRKVRGNARTFMQNSKDTHAGFSEAKE 420

Query: 420 FWLAVNPNYQSEINVEAALNDTESIFTYQKLVLRKEHDLVDADPKLEETADKVPAYVR 479

-2532-

WL VNPNTYQ+INV AL D +SIFPTYQKL+ALRKE DMLN+AD+ LL TADKVFAY R
 Subjct: 421 TWLFPNPNYQ+INVADLADPSIFPTYQKLALRKEQDLVBNADYHLLPTADKVFAYR 480

Query: 480 QTDKERVYLVAHLSDQNSFEFPAVKETIISNTEVQVQLVSSNTLKPMDAPCIEL 534
 Q +E Y+IV N+SD+ Q F A + II+NP+V VL + L+PMDAPC+L L
 Subjct: 481 QLSEETVYIVVNVSDSEQVEATDLAQAQVLIANTVDVTYLETGHLQPMDFACLEL 535

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 2244

A DNA sequence (GBSx2364) was identified in *S. agalactiae* <SEQ ID 6939> which encodes the amino acid sequence <SEQ ID 6940>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have an unclesavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB49738 GB:U21942 UDP-galactose 4-epimerase [Streptococcus mutans]
 Identities = 267/331 (80%), Positives = 306/331 (91%)

Query: 1 MAVLILGGAGYIGSHMVDQLITQGREKVIIVVDNLVTGHRQAVHSDAIFYGDLSDKTEYR 60
 MA+L+LGGAGYIGSHMVD+LI +G+E+V+VVD+LVTGHR AVH A FY+GDL+D+ FM
 Subjct: 1 MAVLILGGAGYIGSHVDRILIRKEEEVTVVDSLVTOHRAAVHFAAKFYQGLADRETFM 60

Query: 61 QVFRNPFDVDAVHFAAFSLVAESMNPFLKYFDNNTAGMIKLLVEMNCEIDKNIVFSSTA 120
 VFRNPFDVDAVHFAAF+SLVAESM+ FLKYFDNNTAGMIKLLVEM+ E +K IVFSSTA
 Subjct: 61 MVFRNPFDVDAVHFAAFSLVAESMNPFLKYFDNNTAGMIKLLVEMNCEIDKNIVFSSTA 120

Query: 121 ATYGIFBQVPILETAPQNPINPYGESKLMETIMKWDQAYGIKFPVALRYFNVAQKEDG 180
 ATYGIP ++PI ET PQ PINPYGESKLMETIMKWD+D+AYGIKFP +RYFNVAQ KEDG
 Subjct: 121 ATYGIPNPIPIKETTPQRPINPYGESKLMETIMKWDRAATGKFPVRYFNVAQKEDG 180

Query: 181 SIGEDHNPETHLLPIILQVAQGVKDKIMIFGDDYNTPDGTINVRDYVHPFDLADHILAVID 240
 SIGEDH NPETHLLPIILQVAQGVK+KIMIFGDDYNTPDGTINVRDYVHPFDLAD H+LA++
 Subjct: 181 SIGEDHNPETHLLPIILQVAQGVKDKIMIFGDDYNTPDGTINVRDYVHPFDLADHLLALN 240

Query: 241 YLRQNSNPNVFNLSSTGFNSNLQMLEAARRITGKRIPAQKAARRPGDPTLIASSEKARQ 300
 YLRQNS N FNLSSTGFNSNLQ+LEAAR++TG++IPA+KAARR GDPTLIASSEKAR+
 Subjct: 241 YLRQNSNPNVFNLSSTGFNSNLQLEAARKVIGQKIPAKKAARRSGDPTLIASSEKAR 300

Query: 301 ILCWEPFNDIDKIISSAWWHSSHHPNGYRD 331
 ++GW+P+FD+I+KII+SAWWHSSH HP GY+D
 Subjct: 301 VVWKEQFDIDKIISSAWWHSSHHPNGYRD 331

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2245

A DNA sequence (GBSx2366) was identified in *S. agalactiae* <SEQ ID 6941> which encodes the amino acid sequence <SEQ ID 6942>. This protein is predicted to be two-component response regulator. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

-2533-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3945 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA006470 GB:AP001516 two-component response regulator [Bacillus halodurans]
 Identities = 71/223 (31%), Positives = 139/223 (61%), Gaps = 7/223 (3%)

10 Query: 3 VLIIEDDPMVVEFIHRYNLEKLNYPQNIYSTASQTQAIAYLNDIKIQVLVDIHIKEGNL 62
 VL+IIEDDPMV+ ++R ++EKL+ F + +TA+ + + +++ L+LIDI + + +GL
 Sbjct: 9 VLIIEDDPMVQVGNRMFVEKLSGFTIVGTTATGREGMVKTRQLQPDLLILDI PMPKQDGI, 68

15 Query: 63 ELLKLLRNQHQNTSEVIVISAANEAMTVKSAFHLGIVDYLKPFPTFERPESIEKFLNHH 122
 +K +R Q+ + ++I ++RAN+ T+K G++DYL+KPPTFER +++++ ++ +
 Sbjct: 69 SFIKIREQYIDVDIIAVTAANDTKITKTLRYGVNDYLVKPFPTFERKLAALTQYEMFR 128

20 Query: 123 TFEAD-KIYQDNIDHFOKIDSGWLBGEVVKLDE--KGLSEITTYQHILDAIQBLEOPPTIQE 179
 + + ++ QD++D K + + +D+ KGL T Q +++ +ELI++P + +E
 Sbjct: 129 KMQKAEISQDSELDENIK--KKQAQANMDDLPGSLHAHTLQQVIERLEELDEPKSASE 184

Query: 180 LAKCSQFSHVSVRKYIAYMEKGLLTSQQIYTKVGRPYKYKL 222
 + + + V+VR+Y+ Y+R G + + Y +GRF +YKL
 25 Sbjct: 185 IGRDVLGARVTVRKYIAYLSEVQVEMDLTSGSIRFIQTYKL 227

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6943> which encodes the amino acid sequence <SEQ ID 6944>. Analysis of this protein sequence reveals the following:

Possible site: 37
 30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.4053 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 123/220 (55%), Positives = 156/220 (70%)

40 Query: 1 MDVLIIEDDPMVVEFIHRYNLEKLNYPQNIYSTASQTQAIAYLNDIKIQVLVDIHIKEGN 60
 M+VLIIEDDPMV+FIHRYNLEKLN F I S + + L D I L+LIDIH I +GN
 Sbjct: 1 MNVLIIEDDPMVDFIHRNLEKLNLFDRITSSDSMKAVQSILTDYADILDLIDHITDGN 60

45 Query: 61 GLELLKLLRNQHQNTSEVIVISAANEAMTVKSAFHLGIVDYLKPFPTFERPESIEKFLNHH 120
 G++ L+ R QH EVI+ISAAN+ + +++ FHGLI+DYLKPFPTFERP+ SI++F+ H
 Sbjct: 61 GIQFLEKRWTKRQIHPCEVILISAANDGNIIRDGFHLGIIDYLIKPFPTFERPQESIQQFVTH 120

Query: 121 YHTFEADKIYQDNIDHFOKIDSGWLBGEVVKLDEKGLSEITTYQHILDAIQBLEOPPTIQEL 180
 ++ Q ID + + S +L EKLSE T+Q I++ I+ +QPPTIQEL
 50 Sbjct: 121 REHLAQQLBQAQIDQLKULTSKKDTKMKQLLEKGLSESTFWIMENKIVFDQPPTIQEL 180

Query: 181 AKCSQFSHVSVRKYIAYMEKGLLTSQQIYTKVGRPYKYKL 220
 A SHVSVRKYIAY+EE L SQI+TKVGRPY+VY
 55 Sbjct: 181 ASACHLSHVSVRKYIAYLEKQLNSQQIFTKVGPRYVY 220

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2246

60 A DNA sequence (GBSx2367) was identified in *S.agalactiae* <SEQ ID 6945> which encodes the amino acid sequence <SEQ ID 6946>. Analysis of this protein sequence reveals the following:

-2534-

```

Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL Likelihood = -8.76      Transmembrane 12 - 28 ( 6 - 34 )
      INTEGRAL Likelihood = -7.43      Transmembrane 178 - 194 ( 173 - 197 )

5

----- Final Results -----
      bacterial membrane --- Certainty=0.4503(Affirmative) < success
      bacterial outside --- Certainty=0.0000(Not Clear) < success
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

A related GBS nucleic acid sequence <SEQ ID 9003> which encodes amino acid sequence <SEQ ID 9004> was also identified. Analysis of this protein sequence reveals the following:

```

15 Lipop: Possible site: -1  Crend: 3
   SCFSLG: 0
   MoG: Length of UR: 27
       Peak Value of UR: 2.99
       Net Charge of CR: 3
   MoG: Discrim Score: 12.92
   GVH: Signal Score (-7.5): -2.57
20   Possible site: 19
   >>> Seems to have an uncleavable N-term signal seq
       Amino Acid Composition: calculated from 1
       ALCM program count: 2 value: -8.76 threshold: 0.0
       INTEGRAL Likelihood = -8.76 Transmembrane 10 - 26 ( 4 - 32)
25       INTEGRAL Likelihood = -7.43 Transmembrane 176 - 192 ( 171 - 195)
       PERIPHERAL Likelihood = 3.18 149
       modified ALCM score: 2.25
       icml HYPID: 7 CFP: 0.450
30 *** Reasoning Step: 3
   ----- Final Results -----
       bacterial membrane --- Certainty=0.4503(Affirmative) < success
       bacterial outside --- Certainty=0.0000(Not Clear) < success
35       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

The protein has homology with the following sequences in the GENPEPT database.

[illegible]

65

-2535-

Query: 358 IYLKELLEQNEFLARLSMLVREPLASFIIGEREKFAKHINLSTELVBIPTKSTVED 417
 Y:K+ Q + + V+ LA F++G++ E+ NL E IP +
 Sbjct: 361 DYIKDIAIQCKSETSEIINDVKSSVLGFLGKGSFIREQGANLDEBCNGVIMAAIDPSV 420

Query: 418 VNNYL-LAHRYINTKILTLN-STTLVSLRLNQNLIETD 456
 ++ ++ ++ IN + + + + + + + N++++ +
 Sbjct: 421 IHELITTIIGNLNLGDAVADMPKKQITMSRFRHNSILDE 461

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6947> which encodes the amino acid sequence <SEQ ID 6948>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood =-10.03 Transmembrane 174 - 190 (170 - 195)

----- Final Results -----
 bacterial membrane --- Certainty=0.5012 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 236/488 (48%), Positives = 337/488 (68%), Gaps = 3/488 (0%)

Query: 3 MKKLSLWAFSLILVTMTICIPSIFFYVTHQSYRMVRVQEKILKNTGYALSRNPQVI 62
 MKK L LNA LSLILY+M + S+FY +H +++ ++ QE +L +TG L+ + +
 Sbjct: 1 MKKPLRLNASLILVSMIVTISLFGYIMLHDTQSIKNGETHLLTSTGMLASHQAIAK 60

Query: 63 QTLKNHDYDQLKQKMLPLSKKSNLDYIVLNLKGIRFTHPDSTKIGKPGQGGDEQAVFK 122
 + L +N + ++ NLDY+V++N+KGR THF+ IGKPPQGGDE+AV
 Sbjct: 61 ELLLNQWAKTATYNTASINLNDYVVMNKGIRLHPNPNKNGKPPQGGDEAVLA 120

Query: 123 GKAIMSTAGSGLKSLRYLIPVVDHQQKQGAIAVGLKLTLDGLQSSIKFESKPLILIS 182
 GK ++S+R+G+LKSRLYL+V+D EQ+GAI+V+G+KLTLT D++ +S + + + LL+ +
 Sbjct: 121 GKKVISTANGTIGKSLRYLVFVFDGDEQIGAI+V+G+KLTTLNVALTSKRNTTISLLLCI 180

Query: 183 LISI+VTSIISYGLKQLNLHPSDIFCHLEERNATLDQIQAAVVIDORHIKRNTPAA 242
 LISL+VTS IS+ LK+QEL L PS+I+Q EERNA LDQI+AAV+V+D I++ N A
 Sbjct: 181 LISL+VTSIFISFLKRLQLHQLPESEITQLFEERNAMLDQIEAAV+VVDKAGILQLCQAG 240

Query: 243 SLLEPKBEGQDLPSGKLLSILPQLQDHPFKKTEQVLHPQGGDYLLSISPTITVKTRRG 302
 L ++ Q + +G L P + + + EQ+ + +DYLL+ISPI VK +RG
 Sbjct: 241 QKLIARKCQLGKPTGNSFNYLPDPFKLSLQGHQLFRYGEEDYLLAISPICVICIDHRG 300

Query: 303 YVYFLRNVTETLFTLDQLAHTTAYASALQACTH+PMKQLHVIYGLADISYDELKYLKE 362
 +++P+R + + TLDQLA+TAYASALQACTH+PMKQLHVIYGL DI YVD+LKIYL
 Sbjct: 301 HIIF+REAVKAIDTLDQLAVTAYASALQACTH+PMKQLHVIYGLDIYVDQKLYLDS 360

Query: 363 LLEQNEFLARLSMLVREPLASFIIGEREKFAKHINLSTELVBIPTKSTVEDVNNYL 422
 +LEP+NE L L+SV+EP LASP+TGE+EK+ E ++L +L RIP +T +N+ L
 Sbjct: 361 LLEPENEILTSLSVLKEPLASFLIGQEKEQYQELNVHLKIDVLSEIPSHQKQNNGL 420

Query: 423 LAHRYINTKILTLNSTTLVSLRLNQNLIETDVCNENEKWL-INHYHQYFNDAFPQQL 481
 +++R+I+T +LT L +LV + QN+LI + + W+ L P+ YPQQL
 Sbjct: 421 MYRPIHTNLLTTLRPKSLVLGICHDQNLIX--SHYPLATONWIDLERVQPIPLPFPQQL 478

Query: 482 LVDSRNTY 489
 L D+ + +
 Sbjct: 479 LTIINSQF 486

SEQ ID 9004 (GBS130d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 8-10; MW 63kDa) and in Figure 184 (lane 4; MW 63kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 11; MW 38kDa) and in Figure 181 (lane 7; MW 38kDa).

-2536-

GBS130d-GST was purified as shown in Figure 237, lane 11. GBS130d-His was purified as shown in Figure 233, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 2247

A DNA sequence (GBSx2368) was identified in *S.galactiae* <SEQ ID 6949> which encodes the amino acid sequence <SEQ ID 6950>. Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence
10  INTEGRAL Likelihood = -11.52 Transmembrane 364 - 380 ( 353 - 386)
    INTEGRAL Likelihood = -9.66 Transmembrane 33 - 49 ( 26 - 57)
    INTEGRAL Likelihood = -7.80 Transmembrane 87 - 103 ( 82 - 105)
    INTEGRAL Likelihood = -6.85 Transmembrane 153 - 169 ( 144 - 174)
15  INTEGRAL Likelihood = -4.41 Transmembrane 301 - 317 ( 300 - 318)
    INTEGRAL Likelihood = -2.81 Transmembrane 216 - 232 ( 212 - 235)
    INTEGRAL Likelihood = -2.39 Transmembrane 120 - 136 ( 120 - 136)
    INTEGRAL Likelihood = -1.65 Transmembrane 57 - 73 ( 56 - 73)
    INTEGRAL Likelihood = -1.17 Transmembrane 428 - 444 ( 428 - 444)
20  INTEGRAL Likelihood = -0.32 Transmembrane 276 - 292 ( 276 - 292)

----- Final Results -----
        bacterial membrane --- Certainty=0.5607 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA18291 GB:U35658 L-malate permease [Streptococcus bovis]
Identities = 329/428 (76%), Positives = 375/428 (86%)
30  Query: 18 DLKAKLFPHIKIGSVLPVTVCLALLILLAGFLQKLPVNMGGFAVILAMGNPLGTIGASI 77
    D + KL + IGSV LPV+ A + IL+ L+LPVNMGGFAVILAMGN LGTIG +I
    Sbjct: 14 DWRNKLTKTIRIGSVTLPLVYLVTASIIIVTALLEQLPVMNKGFAVILAMGNLLGTIGNI 73

    Query: 78 PGFKNFGGPAISLLVPSILVFPNLINKNVLESTNMLMKQANFLYFYIACLVSIGLGN 137
    P K+FGGPAISLLVPSI+VFPNL+N+NVL+ST+LMMKQANFLYFYIACLVSIGLGN
35  Sbjct: 74 PILKHFGGPAISLLVPSIMVFPNLINQNVLSTDLMMKQANFLYFYIACLVCOSILGN 133

    Query: 138 RKQLIQGLLRMIFPMLLGMVCMNVGTFVGVILGLEMRHFLFYIVTPVLGGIGSEILPL 197
    RK+L+QGL+RMI PM LGM+ AM VGI VG +LGL W+H+LFYIVTPVLGGIGSEILPL
40  Sbjct: 134 RKTLVQGLRMIVPMALGMILAMGVGLVTLGLLGLWGHSLFYIVTPVLGGIGSEILPL 193

    Query: 198 SLGYSSITGVASEQLVAQLIPATIIGNFPAILCTALLANRLGEEKPHLSGGQGLVRINKGE 257
    SLGYS+ITG+ SEQVL QLIPATIIGNFFAI+ C+ L+RLGEEK+P LSGGQGL+++ +
    Sbjct: 194 SLGYSAITGLPSELQVQLIPATIIGNFPAIMCSGLLSRIKGRPELSGGQGLTKITIND 253

45  Query: 258 DMSDIIADHSGPIDVKRMGGGVLTACSLFPIGHLQQLTGFGPGVIMTVAARILKYINVI 317
    D+SD + + PIDVK MG GVL AC+LPI G LIQ LTGFGPGVIMIV AA LKY+NV+
    Sbjct: 254 DLSDALESKDAKPIDVKMLGAGVLIACLTPIGTGLQLH+TGFGPGVIMIVVAARILKYINNV 313

50  Query: 318 PRETQNGAKQLYKFISGNFTFPLMAGLGLLYIPLKDVVATLSIQYFIVIVSVVFTVISVG 377
    P+ETQ G+KQLYKFISGNFTFPLM GLG+LYIPLKDVV LS QYE+VVISVFTVI+ G
    Sbjct: 314 PRETQRGSKQLYKFISGNFTFPLMVLGMLYIPLKDVVGNLSQYFVVVISVFTVIATG 373

    Query: 378 FFVSRFLNMNIPVEAGIIISACQSGMGGTGDVAILSTADRNNIMLPPAQVATRLGGAITVITM 437
    FFVSRF+NMNIPVEA I+SAQCSGMGGTGDVAILSTA+RM IMPPAQVATRLGGAITVITM
55  Sbjct: 374 FFVSRFNMNIPVEAAIISACQSGMGGTGDVAILSTANRMTLMPFAQVATRLGGAITVITM 433

    Query: 438 TAILRMLF 445
    TAI RMLF
60  Sbjct: 434 TAI RMLF 441

```

-2537-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6951> which encodes the amino acid sequence <SEQ ID 6952>. Analysis of this protein sequence reveals the following:

Possible site: 48
 5 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -11.89 Transmembrane 361 - 377 (350 - 383)
 INTEGRAL Likelihood = -7.43 Transmembrane 84 - 100 (79 - 102)
 INTEGRAL Likelihood = -6.16 Transmembrane 150 - 166 (137 - 171)
 10 INTEGRAL Likelihood = -4.88 Transmembrane 30 - 46 (24 - 48)
 INTEGRAL Likelihood = -4.35 Transmembrane 299 - 315 (297 - 316)
 INTEGRAL Likelihood = -4.14 Transmembrane 117 - 133 (115 - 134)
 INTEGRAL Likelihood = -3.19 Transmembrane 54 - 70 (51 - 75)
 INTEGRAL Likelihood = -2.92 Transmembrane 425 - 441 (425 - 442)
 INTEGRAL Likelihood = -2.81 Transmembrane 213 - 229 (209 - 232)
 15 INTEGRAL Likelihood = -2.44 Transmembrane 273 - 289 (271 - 290)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5755 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA18291 GB:U35658 L-malate permease [Streptococcus bovis]
 Identities = 344/443 (77%), Positives = 394/443 (88%), Gaps = 6/443 (1%)
 25 Query: 4 ISKKMPQKDLSEHSKAWQNR---RIGSVLPVYLVLATLLVITGWLQQLFVNMLOGFAV 59
 + KK+P +E W+N+ RIGSV LPVYLW A++LVT L+QLFVNMLOGFAV
 Sbjct: 1 MEKKLPATAANETD--WRNKLTTRIGSVLPVYLVTASIIIVTALLQLFVNMLOGFAV 58
 30 Query: 60 ILTGLNLLGTIGATIPGLKHFGGPAILSLVPSILVFPNLLNNVLEATNVMKQANFLY 119
 ILT+GNLLGTIG IP LKHFGGPAILSLVPSI+VFPNLLN NVL++T++LMQANFLY
 Sbjct: 59 ILTMGNLLGTIGKIPILKHFGGPAILSLVPSIMVFPNLLNQVLDSTILMKQANFLY 118
 35 Query: 120 FYIACLVGSSILGMNRKILIQSLFRMIIPMLGMVCMGVGTLVGVILGLQWQHTLFPVV 179
 FYIACLVGSSILGMNRKIL+QSL RMI+PM LGM+ AMGVGTLVG +LGL W+R+LFP+V
 Sbjct: 119 FYIACLVGSSILGMNRKILVQSLNRMIVPMALGMILAMGVGTLVGTLLGLGMWHSFLPYIV 178
 Query: 180 TPVLGGIGEGILFLSLGSAITGVGSEQLVQOLIPATIIIGNFFAILCTALLNRFGKEKP 239
 TPVLGGIGEGILFLSLGSAITG+ SEQLV OLIPATIIIGNFFAI+C+ LL+R GEK P
 40 Sbjct: 179 TPVLGGIGEGILFLSLGSAITGLPSEQLVQOLIPATIIIGNFFAIMCSGLLSRLGKEKP 238
 Query: 240 SYSGQGQLVKIGHSDEMSDALKDNGSALGVKLMGAVGLFACSLFLAGLLQHLQTLDFPGFV 299
 SGGQL+KI +S+D+SDAL+++ +DVLKMGVL AC+LFI GGLLQHLT PPGFV
 45 Sbjct: 239 ELSSGQGLIKITNSDDLDALDEEDKAPIDVKLMGAVGLIACLTPTTGLLQHLQHTGPGFV 298
 Query: 300 LMIILAAFLKYLNVIPQETQNSAKQLYKFISNFTPIIMAGLGLLIYIPLKEVVATLSWQY 359
 LMI++AAFLKYLNV+P+ETQ G+KQLYKFIS NPTFFIM GLG+LYIPLK+VV LSWQY
 Sbjct: 299 LMIVVAAFLKYLNVVPEKTRGSKQLYKFISGNFTPIIMVGLGMLIYIPLKDVVGNLSWQY 358
 50 Query: 360 FIVVISVVLTVVSVGPFVFSRPMNPVEAAIISACQSGMGGTGDAVILSTADRMNLMFPA 419
 F+VVISV TV++ GPFFVSR+NM+PVEAAI+SACQSGMGGTGDAVILSTA+RM LMFPA
 Sbjct: 359 FVVVISVVFVTIATGPFVFSRPMNPVEAAIVSACQSGMGGTGDAVILSTANRMNLMFPA 418
 55 Query: 420 QVATRLGGAITVITMTAILRIIF 442
 QVATRLGGAITVITMTAI R++F
 Sbjct: 419 QVATRLGGAITVITMTAIFRMLF 441

An alignment of the GAS and GBS proteins is shown below.

Identities = 356/419 (84%), Positives = 385/419 (90%)
 60 Query: 27 KIGSVLPVYVCLALILLAGPLQKLPVNMLOGFAVILTMGNLGTIGASIPGFNFGGP 86
 +IGSVLEPVY+ LA LIL+ G+LQ+LPVNMLOGFAVILT+GW LTTIGA+IPG K+FGGP
 Sbjct: 24 RIGSVLPVYLVLATLLVITGWLQQLFVNMLOGFAVILTMGNLGTIGATIPGLKHFGP 83

-2538-

- Query: 87 AILSLVPSILVFPFLINKNVLESTNMIMKQANFLYFYIACLVSGSILGMNRRLIQGLL 146
 AILSLVPSILVFPFLN+H NVLE+TN+IMKQANFLYFYIACLV GSILGMNRK+LIQGL
 Sbjct: 84 AILSLVPSILVFPFLN+HNVLEATNVIMKQANFLYFYIACLVCSILGMNRKILIQGLF 143
- 5 Query: 147 RMIIPMLGMVCMGVGTVGLVGLWRHTLFIYIVTVLAGGIGBSILPLSLGYSITG 206
 RMI PMLLAGVCM VGT VVNLGLH+H+HTLFPY+VTVLAGGIGBSILPLSLGYSITG
 Sbjct: 144 RMIIPMLGMVCMGVGTVGLVGLWRHTLFIYIVTVTVLAGGIGBSILPLSLGYSITG 203
- 10 Query: 207 VASBQLVAQLIPATIIGNFFAILCTALLNRGKKPHLSQQQLVNRKGMHSDIIAD 266
 V SEQVAQLIPATIIGNFFAILCTALLNR GEK P SQQQLV++ EDMSD + D+
 Sbjct: 204 VGSBQLVAQLIPATIIGNFFAILCTALLNRPGSKHPSPYSQQQLVKGHSEDSMDALKN 263
- 15 Query: 267 SGPIDVKMGGGVLTACSLFIPGHLQQLTGFRGPVLMIVAAAILKYINVIPTQNGAK 326
 SG +DVK MG GVLTACSLF I G LQ LT FRGPVLM+ AA LKY+NVIP+STQNGAK
 Sbjct: 264 SGALDKMGAGVLTACSLFPIAGSLQHLTDFFGPVLMITLAAFLKINVIPTQNGAK 323
- 20 Query: 327 QLYKFISSNFTFPMAGLGLLYIPLKDVVATLSIQYFIWISVUVTVISVGFFVSRFLNM 386
 QLYKFIS NFTFPMAGLGLLYIPLK+VVATLS QFIWISVV TV+SVGFFVSRFLNM
 Sbjct: 324 QLYKFISSNFTFPMAGLGLLYIPLK+VVATLSWQYFIWISVVLTVSVGFFVSRFLNM 383
- Query: 387 NPVEAGIISACQSGMGCTGVAILSTADRMHMLPFAQVATRLGGAITVITMTAILRLF 445
 +PVEA IISACQSGMGCTGVAILSTADRMHMLPFAQVATRLGGAITVITMTAILR++F
 Sbjct: 384 SPVEAAIISACQSGMGCTGVAILSTADRMHMLPFAQVATRLGGAITVITMTAILRITP 442

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2248

A DNA sequence (GBSx2369) was identified in *S. agalactiae* <SEQ ID 6953> which encodes the amino acid sequence <SEQ ID 6954>. This protein is predicted to be malic enzyme (mae). Analysis of this protein sequence reveals the following:

- Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.28 Transmembrane 164 - 180 (164 - 181)
- 35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1914 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- 40 The protein has homology with the following sequences in the GENPEPT database.
 >GP:AA07709 GB:U35659 malic enzyme [Streptococcus bovis]
 Identities = 285/386 (73%), Positives = 332/386 (85%), Gaps = 1/386 (0%)
- 45 Query: 2 SENLQQLAINQARENGKLEVISKVKVEKDRDLISIAITPGVASVSSAIARDVELAYELTT 61
 +++ +LAI Q++ + GKKLEV KV +E K DL IATTPGVA+VSSAI B E AVELTT
 Sbjct: 3 TKDVKELAIQAKKFGGKLEVCVKPIETKADLGIAITPGVAVSSAIYRKKERAYELTT 62
- Query: 62 KKNITVAVSDGSAVLGLGIDGPEAAMPVMBGKAALFKRFANVDVPIVLKINDTEEIIISI 121
 KKNITVAV+SDGSAVLGLG+IGPEAAMPVMBGKAALFKRFA VD++P+VL T DTEIISI
 50 Sbjct: 63 KKNITVAVSDGSAVLGLGIDGPEAAMPVMBGKAALFKRFAGVDSIPLWIDTQUTSEIIQT 122
- Query: 122 VKAISPTPGGINLEDISAPRCFPIEQRILIEBDDIPVPHDDQHGTAIVVLAALFNSLKLGVK 181
 VK ++PTPGGINLEDISAPRCFPIEQRIL+ B DIPVPHDDQHGTAIVVLAAL+NSLKL+
 55 Sbjct: 123 VKFLAPTPGGINLEDISAPRCFPIEQRILDELIDIPVPHDDQHGTAIVVLAALFNSLKLGIN 182
- Query: 182 KDIEDIRVVVNGGGSAGLSITRKLSAGAKHVTVDVRFGIINDKRESSLAHHKAIKALT 241
 K IEDI VV+NGGGSAGLSITRK L+AG KH+ +VDR GI+++ D +L PHH IAKLT
 60 Sbjct: 183 KTIEDIHVVINGGGSAGLSITRKFLAAGVGHIIIVDRTGILSETD-TALPHHARIKALT 241
- Query: 242 NREPGSGSLKDALENADVFVIGVSAPEALHARKWISKMDKPIVPMANPIPRIYPDQALKA 301
 NRE ++G L ALE ADVF+GVSA P L ENI +M ++P++FAMANN+PEI+PD+AL A

-2539-

Sbjct: 242 NREHRTGDLATALEGADVFPVGSAPGVLPKPEIQQMBOQVIFAMANVPVEIPDEALAA 301

Query: 302 GAYIVGTGRSDFFMQINNVLAFFGIFRGALDARAKITIVEMQIAAARGIASLIPDEHSLT 361

GAYIVGTGRSDFFMQINNVLAFFGIFRGALDARAK IT:EMQIAAARGIA LIP+ EL+

Sbjct: 302 GAYIVGTGRSDFFMQINNVLAFFGIFRGALDARAKITIVEMQIAAARGIAKLIPDEHSLT 361

Query: 362 TWIIIPAFQNDVADVPVAKSVENAVOK 387

T+IIP+ FQ VA VVA+SV NRV++

Sbjct: 362 TWIIIPAFQNGVAKVAVSVRNVAKE 387

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6955> which encodes the amino acid sequence <SEQ ID 6956>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.44 Transmembrane 164 - 180 (164 - 181)

INTEGRAL Likelihood = -1.75 Transmembrane 94 - 110 (94 - 110)

----- Final Results -----

bacterial membrane --- Certainty=0.1977 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AB07709 GB:U35659 malic enzyme [Streptococcus bovis]

Identities = 289/379 (76%), Positives = 334/379 (88%), Gaps = 1/379 (0%)

Query: 7 QALEQAKITFGKLEVPQKVDIKTKHDLIAITPGVASVSSAIAKDTLAYDLATKIGTV 66

+IA+EQAK FQKLEVPK V+TK DL IAITPGVA+VSSAI + K AY+LITKIGTV

Sbjct: 8 ELAIEQAKKFGKLEVPCKVPIETKADLGIATPGVAAVSSAIYKKERAYELTKKNTV 67

Query: 67 AVISDGTAVLGLGDIQPEAMPVMSGAALFKRFAGVDPIVLDTKTEIIISIVKALA 126

AVISGQ+AVLGLG+IGPEAMPVMSGAALFK RFAGVD+IP+VLDT+DTEII VK LA

Sbjct: 68 AVISDGSVAVLGLGDIQPEAMPVMSGAALFKRFAGVDSIPVLDTQDTEIIITVKFLA 127

Query: 127 PTFGGINLEDISAPRCFRIEQRLLKECHIPVFHDDQHGTAIVVLAARFNSLKLAKSLDE 186

PTFGGINLEDISAPRCFRIEQRLL E IPVFHDDQHGTAIVVLAAR+NS+KL+ K +++

Sbjct: 128 PTFGGINLEDISAPRCFRIEQRLLDELDIPVFHDDQHGTAIVVLAALYNSLKLKINKIED 187

Query: 187 VSIUVN3GGSAGLSITRKLAAAGATKVTVDKRGIIINQESAQLAPHHLDIAKVTNREPK 246

+ +V+NGGGSAGLSITRKL LAAG + +VD+ GI++E + A L PHH +IAK+TIRE +

Sbjct: 188 IHVINGGGSAGLSITRKLAAAGVXIIIVDRTOILSETDTA-LPPHAEIAKLVNREHR 246

Query: 247 SGTLEDALEGADIPGVSAPOVLKAEWISKMAARPVIFAMNPPIPIYVDEALEAGVIV 306

+G L ALEGAD+P+GVSAPOVLK EMI +M +PVIFAMNP+PEI+PDEAL AGVIV

Sbjct: 247 TSDLATALEGADVFPVGSAPGVLPKPEIQQMBOQVIFAMANVPVEIPDEALAAAGVIV 306

Query: 307 GTRSDFFMQINNVLAFFGIFRGALDARAKITIVEMQIAAARGIASLIPDEHSLTNIIP 366

GTRSDFFMQINNVLAFFGIFRGALDARAK IT+EMQIAAARGIA L+PD+ L+ TNIIP

Sbjct: 307 GTRSDFFMQINNVLAFFGIFRGALDARAKITIVEMQIAAARGIAKLIPDEHSLTNIIP 366

Query: 367 DAFKGVAVIAKSVRSV 385

D +KGVAA+VA+SVR+ V

Sbjct: 367 DPFQGVAVIAKSVRSV 385

An alignment of the GAS and GBS proteins is shown below.

Identities = 306/387 (79%), Positives = 349/387 (90%)

Query: 1 MSENGLQALINQARENGKLEVIISKVVEDKRDLSIAITPGVASVSSAIAEDVELAYELT 60

M IGLQA+ QA+ GKLEVP KV ++ K DLSIAITPGVASVSSAIA+D LAY+IT

Sbjct: 1 MKNQLGQALEQAKITFGKLEVPQKVDIKTKHDLIAITPGVASVSSAIAKDKLAYDLT 60

Query: 61 TKKNTVAVVSDGSVAVLGLGDIQPEAMPVMSGAALFKRFANVDVPIVLKNTDTEIIS 120

TKGNTVAV+SDG+AVLGLGDIQPEAMPVMSGAALFK FAVD+PIVL T DTEIIS

Sbjct: 61 TKGNTVAVISDGTAVLGLGDIQPEAMPVMSGAALFKRFANVDVPIVLDTKTEIIS 120

-2540-

```

Query: 121 IVKAISPTFGGINLEISAPRCFIEQRILKECTIIPVFDHQQHGTATVVLAALFNSGLKLV 180
      IVKA++PTFGGINLEISAPRCFIEQRILI+EC 1PVFDHQQHGTATVVLAALFNSGLK+
5  Sbjct: 121 IVKALAPTFGGINLEISAPRCFIEQRILKECHIFVFDHQQHGTATVVLAALFNSGLLL 180

Query: 181 KKDIEDIRVVVNGGGGAGLSITRKLISAGAKHVTVDVRFGIINOKDRSLAPHHIAIKL 240
      KK +++++VVVNGGGGAGLSITRKL+AGA VTVD+FGIIN+++ LAPHH IAK+
20 Sbjct: 181 KKSLEDESVVNGGGGAGLSITRKLAAAGATKVTVDKFGIINBCEAAQLAPHHILDAKV 240

Query: 241 TNREFSGSLEDALENADVFIGVSAPEALHAEWISGMADKPIVPMANPIPEIYPDQALK 300
      TNREF+SG+LEDALE AD+FIGVSAP L AEWISMA +P++FMANPIPEIYPD+AL+
30 Sbjct: 241 TNREFKSGTLEDALEGADIFIGVSAPGLKAEWISMAARPVIFMANPIPEIYPDLEAL 300

Query: 301 AGAIVGTGRSDFPNQINNVLAFFGIFRGALDARAKTITVEMQIAAARGIASLIPSEKLS 360
      AGAIVGTGRSDFPNQINNVLAFFGIFRGALDARAKTITVEMQIAAARGIASL+P++ 16
15 Sbjct: 301 AGAIVGTGRSDFPNQINNVLAFFGIFRGALDARAKTITVEMQIAAARGIASLIPDNL 360

Query: 361 TTHIIPNAFQNVADVAKSVSNVAVK 387
      TT+IIP+AF+ VA++VAKSV + V K
20 Sbjct: 361 TTHIIPDAFKGVAEIVAKSVRSVVLK 387

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2249

- 25 A DNA sequence (GBSx2370) was identified in *S. agalactiae* <SEQ ID 6957> which encodes the amino acid sequence <SEQ ID 6958>. This protein is predicted to be Bta. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence
30  INTEGRAL    Likelihood = -2.02    Transmembrane    29 - 45 ( 29 - 45)

----- Final Results -----
      bacterial membrane --- Certainty=0.1808 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
35  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
Identities = 35/112 (31%), Positives = 63/112 (56%)
40  Query: 1  MYSFELLALMTLITAAIEDKIDNQDFVLFIGRISCFCHLFPVKIHFVADEDEFELF 60
      M F + + + + T + + + D + FIGR +CF+ C F + V E + + +
Sbjct: 1  MEQFLDNKIDLEVTTVVRAQALDKKETATFFIGRKTFCYCRKFAGTSGVVAETKAHIT 60

45  Query: 61  HLDSEDFDHMTANKSFRNKYDIPTVGLMVVKNQTIKVKCDKSMITKEIREF 112
      ++SE+ + FR++Y IPTVPG + + +G I V+CDK M+ +EI++F
Sbjct: 61  FINSEASQINDLQAFRRRYGIPTVGFVHITDQINVRCDSSMSAQEIKDF 112

```

- 50 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6959> which encodes the amino acid sequence <SEQ ID 6960>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55  bacterial cytoplasm --- Certainty=0.0900 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-2541-

Identities = 39/111 (35%), Positives = 66/111 (59%)

Query: 3 SFELLATMTLITAAEIEDKIDSNQDFVLFIGRISCPCHLFLVFKIVEADEDEFELPHL 62
 +FEE+AA + AE+ I E +D ++F+GR SCP+C F PK+ +VA +++ E++ +
 5 Sbjct: 11 TFEEIVANFIPSSVARVTSALASGKMIVFLGRSSCPYCRFPAPKLAQVATNQKEVYFV 70

Query: 63 DSEDFDHWTKANKEFRNKYDIPVPLMVVKNGTIKVKCDKMTKEBIREFI 113
 DSE+ FR Y +TVP L+V + + CDS +T ++I F+
 10 Sbjct: 71 DSENAADAELAAPRENYQLVTVFALLVSYDQHQRACVDSLITFDILAFL 121

SEQ ID 6958 (GBS427) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 5; MW 16.2kDa).

GBS427-His was purified as shown in Figure 214, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2250

A DNA sequence (GBSx2371) was identified in *S. agalactiae* <SEQ ID 6961> which encodes the amino acid sequence <SEQ ID 6962>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.75 Transmembrane 2 - 18 (1 - 21)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9437> which encodes amino acid sequence <SEQ ID 9438> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA11328 GB:D78257 ORF11 [Enterococcus faecalis]
 Identities = 36/80 (45%), Positives = 58/80 (72%)
 35 Query: 1 MSLPIMILVMVGMFFMQRQKKQQAQERQQLNAVQKGEIVTIGGLPGVVDEVNTEAQ 60
 M L +IML+V+V M F++ R QKKQ +ERQ LN +Q GD +VTIGGL GV+ E++++ +
 Sbjct: 1 MKMLMLLLVIVAMYPFLFRTQKKQKQERQDPLNLLQPGDAVVITIGGLGVISEISSDKK 60

Query: 61 RMVLVDVGVYLTFELAAIKS 80
 ++ LD +G + F+ +I++
 40 Sbjct: 61 KVTLOCEGAFFDFDQQSIRT 80

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6963> which encodes the amino acid sequence <SEQ ID 6964>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -6.10 Transmembrane 3 - 19 (1 - 22)
 INTEGRAL Likelihood = -3.03 Transmembrane 63 - 79 (63 - 79)
 ----- Final Results -----
 50 bacterial membrane --- Certainty=0.3442(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-2542-

>GP:BA11328 GB:D78257 ORF11 [Enterococcus faecalis]
Identities = 29/75 (38%), Positives = 52/75 (68%)

5 Query: 6 ILMFVFMGLIWMQRQKKQAQERQKQLNAIEKGDEVVTIGGMFAIVDEVDTAKGIVL 65
++M +V++ + ++ R QKKQ +ERQ+ LN ++ GD VVTIGG+ ++ E+ + KK+ L
Sbjct: 5 LILMLVIVAMFYFLFRFQKQKKERQQLNNLQFGDAVVTIGGLHGVISISSDKKRVIL 64

10 Query: 66 LDVDGVLTTFELLAIK 80
D +G F F+ +I+
Sbjct: 65 DCEGAFFDFDQGSIR 79

An alignment of the GAS and GBS proteins is shown below.

Identities = 63/90 (70%), Positives = 80/90 (88%)

15 Query: 4 PIIMLVVMGMFFMQRQKKQAQERQKQLNAVQKGSIVTIGLFGVVDEVNTEAQRNV 63
PI+M VVM+G+++FMQRQKKQAQERQ QANA++KGDE+VTIGG+F +VDEV+T A++V
Sbjct: 5 PIIMLVVMGLIWMQRQKKQAQERQKQLNAIEKGDEVVTIGGMFAIVDEVDTAKKIV 64

20 Query: 64 LDVDGVYLTTFELAAIKSVSKAATPTEPVE 93
LDVDGV+LTFEL AIK +V+KA T T VE
Sbjct: 65 LDVDGVLTTFELAIKRIVTATTE+TLVE 94

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 2251

A DNA sequence (GBSx2372) was identified in *S. agalactiae* <SEQ ID 6965> which encodes the amino acid sequence <SEQ ID 6966>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have an uncleavable N-term signal seq

30 ----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 2252

A DNA sequence (GBSx2373) was identified in *S. agalactiae* <SEQ ID 6967> which encodes the amino acid sequence <SEQ ID 6968>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -1.38 Transmembrane 164 - 180 (164 - 180)

----- Final Results -----
bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB61731 GB:AL13220 putative oxidoreductase. [Streptomyces

-2543-

- coellicolor A3(2)]
 Identities = 72/216 (33%), Positives = 120/216 (55%), Gaps = 1/216 (0%)
- 5 Query: 14 AQLAEARQKLYSVAMRTYDKLEPATKYGIQKVVDHIDQVPEDPEVDIIYISPTNTHI 73
 A ++ ++ ++ +VA+KT FA ++GI + Y ++ D +VD++Y++TPH+ H
 Sbjct: 25 ADLVDLPDAEVVAVASAKTEASAKTFAERRGIPRAYGGMETLARDEDVDVVVYVATPHSAHR 84
- 10 Query: 74 SFLKCALANGKHVLCEKSTITANSTRLKRAIDLAETNHVLAERMTIFPMPTVRLQKLTAD 133
 + L G+VLCCK TIN+ E E +IA N V L EAM ++ P+ R+LK LV
 Sbjct: 85 TAAGLCLEAGRNVLCEKPPITLAREAAELVALARENGVFLMEAMMYCNPLVRLKELVA 144
- Query: 134 SKLGLPLAKICQNGFSYKYEIMNRFPFSRLAGGALLDIGVALSCIRWPNSEAPHNITS 193
 G +G ++ +Q +FG + +R G3ALLD+GVY +S + + E P ++ +
 Sbjct: 145 DGAIGEVRSLQADPLAGPFPFAHRLRDPAQ33CALLDLGVYVPSFAQLLLGE -PTDVAA 203
- 15 Query: 194 QVTFAPTGVDEQVGILLINPANEMATVSLSLHAKQP 229
 + + + GVD Q G LL+ + +A++ S+ P
 Sbjct: 204 RAVLSEEGVDLQTGALLSYGNDALASIHCSITGTFP 239
- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2253

- A DNA sequence (GBSx2374) was identified in *S. agalactiae* <SEQ ID 6969> which encodes the amino acid sequence <SEQ ID 6970>. This protein is predicted to be surface protein Rib. Analysis of this protein
- 25 sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

- 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4957(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S. pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2254

- A DNA sequence (GBSx2375) was identified in *S. agalactiae* <SEQ ID 6971> which encodes the amino acid sequence <SEQ ID 6972>. This protein is predicted to be surface protein Rib. Analysis of this protein
- 40 sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S. pyogenes*.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2544-

Example 2255

A DNA sequence (GBSx2376) was identified in *S.agalactiae* <SEQ ID 6973> which encodes the amino acid sequence <SEQ ID 6974>. This protein is predicted to be a host cell surface-exposed lipoprotein. Analysis of this protein sequence reveals the following:

```

5  Possible site: 38
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -7.75    Transmembrane    9 - 25 ( 5 - 28)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9005> which encodes amino acid sequence <SEQ ID 9006>
15 was also identified. Analysis of this protein sequence reveals the following:

      Lipop: Possible site: -1    Crend: 3
      SRCFLG: 0
      MoG: Length of UR: 24
          Peak Value of UR: 2.84
          Net Charge of CR: 2
20  MoG: Discrim Score: 10.29
      GVH: Signal Score (-7.5): -4.34
          Possible site: 34
          >>> Seems to have an uncleavable N-term signal seq
      Amino Acid Composition: calculated from 1
      ALON program count: 1 value: -7.75 threshold: 0.0
          INTEGRAL    Likelihood = -7.75    Transmembrane    5 - 21 ( 1 - 24)
          PERIPHERAL Likelihood = 13.31      86
30  modified ALON score: 2.05
      icml HYPID: 7    CFP: 0.410

      *** Reasoning Step: 3

35  ----- Final Results -----
      bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40  >GP:AAC03455 GB:AF020798 putative host cell surface-exposed
      lipoprotein [Streptococcus thermophilus bacteriophage TP-J34]
      Identities = 40/102 (39%), Positives = 63/102 (61%), Gaps = 10/102 (9%)

45  Query: 101 KNALISAKIYKTMNLKSKQSIPEQLYSESPDKATHSDKPTTKRESQYATIDHLKVDKFNAL 160
      + A+ AK Y+ T++SK+ + QL S      DK+++ S YA+++ +D+ + AL
      Sbjct: 51 RTAVSKAKQYASTVEMSKELRLSQLVS-----FDKYSQDASDYAVENSIGDYNKQAL 102

      Query: 161 ETAKYSQSSSSLSKREIYKQLTSTLGDKPTNDEAQYAVDHLK 202
      E AK YQ + S+ + I QL S      DKPT +EA YAV +LK
50  Sbjct: 103 EKAKYQVITLSMSPDAIRDLQLVSP--DKFTQEADYAVANLK 142
      Identities = 40/112 (35%), Positives = 64/112 (56%), Gaps = 9/112 (8%)

      Query: 41 KKAKIK+KNTQKKIVKQKAREYAKSGHMSKDSITEKLKDSKKYQBEDINFVIMNLKVDYK 100
      + ++ K K + V KA++YA + HMSK+ + +L K Y Q+ ++ + N +DY
55  Sbjct: 40 QSSSEKVPKBYRTAVSKAKQYASTVEMSKELRLSQLVSFDK-YSDASDYAVENSIGDYN 98

      Query: 101 KNALISAKIYKTMNLKSKQSIPEQLYSESPDKATHSDKPTTKRESQYATIDHLK 152
      K AL AK Y T++S +I +QL S      DKPT+EE+ YA++LK
60  Sbjct: 99 KQALEKAKYQVITLSMSPDAIRDLQLVSP-----FDKFTQEADYAVANLK 142

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-2545-

SEQ ID 9006 (GBS122) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 6; MW 21.9kDa).

GBS122-His was purified as shown in Figure 202, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2256

A DNA sequence (GBSx2377) was identified in *S.agalactiae* <SEQ ID 6975> which encodes the amino acid sequence <SEQ ID 6976>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2830 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GF:CA90833 GB:A7250837 hypothetical protein [Streptococcus dysgalactiae]
Identities = 91/96 (94%), Positives = 93/96 (96%)

Query: 1 MSRKVRHHFTDDFKQQIVELYNVGRKRSSLIKVYELTPSTFDKMWVQAQTTGSPKSIDNL 60
      MSRK+RRHFTDDFKQQIVELYN GRKRSSLIK YELTPSTFDKMWVQAQTTGSPK+DNL
Sbjct: 1 MSRKIRRHFTDDFKQQIVELYNAGRKRSSLIKYELTPSTFDKMWVQAQTTGSPKSDNL 60

Query: 61 TDEQRELIELRKHNKELEMQLDLKQAQVIMAQKQK 96
      TDEQRELIELRK NKELEMQLDLKQAQVIMAQKQK
Sbjct: 61 TDEQRELIELRKHNKELEMQLDLKQAQVIMAQKQK 96

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2257

A DNA sequence (GBSx2378) was identified in *S.agalactiae* <SEQ ID 6977> which encodes the amino acid sequence <SEQ ID 6978>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2618 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9915> which encodes amino acid sequence <SEQ ID 9916> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9903> which encodes amino acid sequence <SEQ ID 9904> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2546-

>GP:CAB90834 GB:A7250837 putative transposase [Streptococcus dysgalactiae]
Identities = 243/259 (93%), Positives = 250/259 (95%)

5 Query: 1 MCRWLNMPSSYYTQAVSVSETEFEETIKRIFLDGSESYSGRKIKICLANNEGITLSRRR 60
MCRWLN+P SSYYT+AVR VSE E KE+IK IFL+S++RYGSKKIKICLANNEGITLSRRR
Sbjct: 1 MCRWLNIPRSSYYTQAVEPVSEAELESIKAFLESKARYSGRKIKICLANNEGITLSRRR 60

10 Query: 61 IRRIMKRLNLVSVTQKATFKPHSRGKNEAPIFNHLDQPKQERPLQAVTDLTYVRVGNR 120
IRRIKRLNLVSVTQKATFKPHSRGKNEAPIFNHLDQPK ERPLQAVTDLTYVRVGNR
Sbjct: 61 IRRIMKRLNLVSVTQKATFKPHSRGKNEAPIFNHLDQPKQERPLQAVTDLTYVRVGNR 120

15 Query: 121 WAYVCLIIDLYNREIIGLSLGHKTAELVKQAIQSIPYALTKVRMPHSDRGKEPDNQID 180
WAYVCLIIDLYNREIIGLSLGHKTAELVKQAIQSIFY LTKVRMPHSDRGKEPDNQID
Sbjct: 121 WAYVCLIIDLYNREIIGLSLGHKTAELVKQAIQSIPYALTKVRMPHSDRGKEPDNQID 180

20 Query: 181 EILEAFGITRSLSQAGCPYDनावेष्ट्याफकिरव्युत्तपल्लेालकतुधिवम्वन्य 240
EILEAFGITRSLSQAGCPYDनावेष्ट्याफकिरव्युत्तपल्लेालकतुधिवम्वन्य
Sbjct: 181 EILEAFGITRSLSQAGCPYDनावेष्ट्याफकिरव्युत्तपल्लेालकतुधिवम्वन्य 240

25 Query: 241 HRIHGSLSNYQTWTKRLIA 259
HRIHGSLSNYQTWTKRLIA
Sbjct: 241 HRIHGSLSNYQTWTKRLIA 259

There is also homology to SEQ ID 32.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2258

A DNA sequence (GBSx2379) was identified in *S.galactiae* <SEQ ID 6979> which encodes the amino acid sequence <SEQ ID 6980>. This protein is predicted to be pXO1-128. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.3684(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AD32432 GB:AF065404 pXO1-128 [Bacillus anthracis]
Identities = 45/69 (65%), Positives = 52/69 (75%)

45 Query: 17 MKKAGKSNRVIMKTLGIKNNSQITWMMKWRNKRILYRPHQGVQKQPTYGKGLHLSVEVQ 76
MKK SNR IME LGIKH SQI TWMMKWR ++ YRP Q VGRQY+YSGK + LSR+EQ
Sbjct: 1 MKKESYSNRTIMEHLGIKNVQGIKTMKWRNRTQTRFQQYFVGKQSYSGKSPKLSLSLQ 60

Query: 77 LQLQVDDLK 85
L+L+ LK
Sbjct: 61 LRLENHKLK 69

50

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2547-

Example 2259

A DNA sequence (GBSx2380) was identified in *S.agalactiae* <SEQ ID 6981> which encodes the amino acid sequence <SEQ ID 6982>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

```

5      Possible site: 25
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2260

A DNA sequence (GBSx2382) was identified in *S.agalactiae* <SEQ ID 6985> which encodes the amino acid sequence <SEQ ID 6986>. This protein is predicted to be Lmb. Analysis of this protein sequence

```

20 reveals the following:

      Possible site: 18
      >>> May be a lipoprotein

      ----- Final Results -----
25      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1595> which encodes the amino acid sequence <SEQ ID 1596>. Analysis of this protein sequence reveals the following:

```

      Possible site: 18
      >>> May be a lipoprotein

      ----- Final Results -----
35      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40      Identities = 302/306 (98%), Positives = 303/306 (98%)

Query: 1   MKKVFFIAMMVSVLWITAGCKSANPKQPTQGMSVVTSPYPMYAMTKEVSGLDNDVRMIQ 60
      MKK  FFLAMMVSVLWITAGCKSANPKQPTQGMSVVTSPYPMYAMTKEVSGLDNDVRMIQ
45      Sbjct: 1   MKKGFFLAMMVSVLWITAGCKSANPKQPTQGMSVVTSPYPMYAMTKEVSGLDNDVRMIQ 60

Query: 61  SGAGTHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKSKVNVFEASKPLTLDR 120
      SGAGTHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKSKVNVFEASKPLTLDR
46      Sbjct: 61  SGAGTHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKSKVNVFEASKPLTLDR 120

Query: 121 VKGLEDMEVTQGDIPATLYDPHWTDPVLAGEAVNIAKELGLDPIGHKDSYTKKAKGFK 180
      VKGLEDMEVTQGDIPATLYDPHWTDPVLAGEAVNIAKELGLDPIGHKDSYTKKAKGFK
50      Sbjct: 121 VKGLEDMEVTQGDIPATLYDPHWTDPVLAGEAVNIAKELGLDPIGHKDSYTKKAKGFK 180

Query: 181 KEAQUTEESYTKKFKKVRSKTFVTQHTAFSLAKRPLQKLGISGISEPEQESPRQLKEI 240

```

-2548-

```

      KEABQLTSEYTKPKKVRSKTFFVQHTAFSYLAKRFLKQLIGISGSPQEPSPRLQKEI
Sbjct: 181 KEABQLTSEYTKPKKVRSKTFFVQHTAFSYLAKRFLKQLIGISGSPQEPSPRLQKEI 240
Query: 241 QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLRAAPSGNKITYLENLRANLEV 300
      QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLRAAPSGNKITYLENLRANLEV
Sbjct: 241 QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLRAAPSGNKITYLENLRANLEV 300
Query: 301 LYQQLK 306
      LYQQLK
Sbjct: 301 LYQQLK 306

```

There is also homology to SEQ ID 4.

SEQ ID 6986 (GBS189) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 2; MW 35.2kDa).

- 15 The GBS189-His fusion product was purified (Figure 204, lane 7) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 248A), FACS (Figure 248B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoreactive on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 2261

A DNA sequence (GBSx2383) was identified in *S. agalactiae* <SEQ ID 6987> which encodes the amino acid sequence <SEQ ID 6988>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4656 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AA541455 GB:U34956 phosphoribosylformylglycinamide synthase
[Mycobacterium tuberculosis]
35 Identites = 73/237 (30%), Positives = 112/237 (46%), Gaps = 25/237 (10%)
Query: 43 GAGGVCVVAIGELAD----GLRIDLDKVLKYLQGLAKFRIALSRQRRMSVVVGGSDVDAP 98
      G G+ A ELA G+ I LD VPL+ + + E+ SESQERM VV P +VDAP
Sbjct: 282 GAGGLSCATSELASAGDGGMTTQLDGVPLRAKEMTPARVLCSSESQRMCAVSPKNDAP 341
40 Query: 99 IAAENKENIDAVVVATVTEKNLVMTVNGETIVDLERCFLDTNG-----VEVVVDKVV 152
      +A C K + A V+ VT+ L +TW+GET+VD+ + G V +
Sbjct: 342 LAVCRKVEVLATVIGEVTDGRLQITWGHSTVVDVVEPRTVAHIGEPVYQRPVAREDTQDAL 401
45 Query: 153 DKDLTVPEARTTSARTLSADMLKVLSDLNHASQKQLQTI FDSVGRSTV--NHPIGGRYQ 210
      + D + +R + + L A +L +L + S+ + +D V +TV H GG +
Sbjct: 402 NADRSAKLSRPVTGDELRAITLALLGSPHLCRAFITBQYIRYVRGNTVLAHAGGNLR 461
50 Query: 211 ITPTSSSVQKLVPVQYGVTTTASVMAQQYNFYIAEWSYHGAAYAVIETARIVATGA 267
      I ES+ + + V + +++ PY GA A+ EA + TGA
Sbjct: 462 I--DESTGRGLAVSTDAISGRYTL-----DPYAGAQLALAEAYRNVAVTGA 505

```

There is also homology to SEQ ID 982.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2262

A DNA sequence (GBSx2384) was identified in *S. agalactiae* <SEQ ID 6989> which encodes the amino acid sequence <SEQ ID 6990>. This protein is predicted to be 30S ribosomal protein S11 (rpsK). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0598 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9281> which encodes amino acid sequence <SEQ ID 9282> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10919> which encodes amino acid sequence <SEQ ID 10920> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:GAB11918 GB:Z99104 ribosomal protein S11 (BS11) (Bacillus subtilis)
Identities = 81/92 (88%), Positives = 87/92 (94%)

Query: 2 HGNALWSSAGALGFKGSRKSTPPAAQMAAFAAANKSAQEHGLKTVEVTVKGPQSGRESAI 61
HGNAAWSSAGALGFPKSRKSTPPAAQMAAFAAANKSAQEHGLKTVEVTVKGPQSGRESAI
Sbjct: 40 HGNALWSSAGALGFKGSRKSTPPAAQMAAFAAANKSAQEHGLKTVEVTVKGPQSGRESAI 99

Query: 62 RALAAAGLEVTIAIRDVTVPVPHNGARPPKRRRV 93
RALAAGLEVTIAIRDVTVPVPHNGARPPKRRRV
Sbjct: 100 RALAAAGLEVTIAIRDVTVPVPHNGARPPKRRRV 131

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6991> which encodes the amino acid sequence <SEQ ID 6992>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0945 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 92/93 (98%), Positives = 93/93 (99%)

Query: 1 HGNALWSSAGALGFKGSRKSTPPAAQMAAFAAANKSAQEHGLKTVEVTVKGPQSGRESA 60
HGNALWSSAGALGFKGSRKSTPPAAQMAAFAAANKSAQEHGLKTVEVTVKGPQSGRESA
Sbjct: 35 HGNALWSSAGALGFKGSRKSTPPAAQMAAFAAANKSAQEHGLKTVEVTVKGPQSGRESA 94

Query: 61 IRALAAAGLEVTIAIRDVTVPVPHNGARPPKRRRV 93
IRALAAAGLEVTIAIRDVTVPVPHNGARPPKRRRV
Sbjct: 95 IRALAAAGLEVTIAIRDVTVPVPHNGARPPKRRRV 127

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2550-

Example 2263

A DNA sequence (GBSx2385) was identified in *S.agalactiae* <SEQ ID 6993> which encodes the amino acid sequence <SEQ ID 6994>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2551 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB03881 GB:AP001507 DNA-directed RNA polymerase alpha subunit
[Bacillus halodurans]
Identities = 190/314 (60%), Positives = 249/314 (78%), Gaps = 2/314 (0%)

Query: 1 MIEFEKPIITTKIDENKD--YGRFVIEPLERGYGTTLGNLSRRVLLSSLPAAVTSIKIDG 58
MIE EKP+I I+ +D YG+V+EPLERGYGTTLGNLSRR+LSSLPAAVTS++IDG
Sbjct: 1 MIEIEKPVIEITEISDAKYGKRVVEPLERGYGTTLGNLSRRILLSSLPAAVTSVQIDG 60

Query: 59 VLHEFDITPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVGPAPETAGDILTDSIDIEI 118
VLHEF TI GV EDV I+LN+K LA+K Y +++K +E+D +G +TAGD+ DSD+++
Sbjct: 61 VLHEFSTIESGVVEDVTIILNLKQALKIYSDEDKTLEIDTQCGSVVTAGDLTHDSVDV 120

Query: 119 VNPDEYLPTIABGHSKIAKMTVAIQNGYVFAEGNKKDDAPVGLTAVDSIYTPVKKNYQV 178
+NED ++ T+ G L+ +T +RGYVFAEGNK D+ +G + +DSIYTPV +VNYQV
Sbjct: 121 INPDLIATLTGAHLRMRITAKRGYVFAEGNKSDELAIGVIFDSIYTPVSRVNYQV 180

Query: 179 EPARVGNDSGDFDKLTIEIMNGTIIPEDALGLSARVLEHNLNLTDLTAVAKATEVMKET 238
E RVG +DKT+++ T+G+I P+*A+ L A++L EHLN+F LT+ A+ E+M E
Sbjct: 181 ENTRVGQVTRYDKLTLDWTDGSIREFEAVSLGAKILTEHLNIFVGLDQAQMEIMVEK 240

Query: 239 EKVNDEKLVDRITIEELDSVRSYNCLKAGINTVFDLTETKEPEMMKVRNLGRKSLEEVK 298
E+ EKVL+ TIEELDSVRSYNCLKAGINTV +LT+KTE +NMKVRNLGRKSLEEV+
Sbjct: 241 EEDQKEKVLWMTIEELDSVRSYNCLKAGINTVQELTQKTEEDNMKVRNLGRKSLEEVQ 300

Query: 299 IKLADLGLGLKNDK 312
KL +LGLGL+ ++
Sbjct: 301 EKLGLGLGLRKEE 314

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6995> which encodes the amino acid sequence <SEQ ID 6996>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2551 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 305/312 (97%), Positives = 311/312 (98%)

Query: 1 MIEFEKPIITTKIDENKDYGRFVIEPLERGYGTTLGNLSRRVLLSSLPAAVTSIKIDGVL 60
MIEFEKPIITTKIDENKDYGRFVIEPLERGYGTTLGNLSRRVLLSSLPAAVTSIKIDGVL
Sbjct: 1 MIEFEKPIITTKIDENKDYGRFVIEPLERGYGTTLGNLSRRVLLSSLPAAVTSIKIDGVL 60

Query: 61 HEFDITPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVGPAPETAGDILTDSIDIEIVN 120
HEFDITPGVREDVMQIILNVKGLAVKSYVEDEKIIEL+VGPAP+TAGDILTDSIE+VN
Sbjct: 61 HEFDITPGVREDVMQIILNVKGLAVKSYVEDEKIIEL+VGPAP+TAGDILTDSIDIEIVN 120

```


-2551-

Query: 121 PDHYLPTIABGHSLKATMTIVAKNGYVPAEGNKDDAPVGTI.LAVDSIYTPVKKNVYQVEP 180
 PDHYLPTIABGHSL+ATMTIVAK RGYVPAEGNKDDAPVGTI.LAVDSIYTPVKKNVYQVEP
 Sbjct: 121 PDHYLPTIABGHSLKATMTIVAKNGYVPAEGNKDDAPVGTI.LAVDSIYTPVKKNVYQVEP 180

5 Query: 181 ARVGSNDGFDKLTIEIMINGTIIIPEDALGLSARVLIEHLNLPDILTVEAKATEVMKETER 240
 ARVGSNDGFDKLTIEIMINGTIIIPEDALGLSARVLIEHLNLPDILTVEAKATEVMKETER
 Sbjct: 181 ARVGSNDGFDKLTIEIMINGTIIIPEDALGLSARVLIEHLNLPDILTVEAKATEVMKETER 240

10 Query: 241 VNDEKVLDRITIEELDSVRSYNCLAKRAGINTVFDLTEKTEPEMMKVRNLGRKSLEEVKIK 300
 VNDEKVLDRITIEELDSVRSYNCLAKRAGINTVFDLTEK+EFEMMKVRNLGRKSLEEVK+K
 Sbjct: 241 VNDEKVLDRITIEELDSVRSYNCLAKRAGINTVFDLTEKSEPEMMKVRNLGRKSLEEVKVR 300

Query: 301 LADLGLGLKNDK 312
 LADLGLGLKNDK
 15 Sbjct: 301 LADLGLGLKNDK 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2264

20 A DNA sequence (GBSx2386) was identified in *S. agalactiae* <SEQ ID 6997> which encodes the amino acid sequence <SEQ ID 6998>. This protein is predicted to be 50S ribosomal protein L17 (rplQ). Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1609 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP: CAB11920 GB: Z99104 ribosomal protein L17 (B15) [Bacillus subtilis]
 Identities = 95/128 (74%), Positives = 105/128 (81%), Gaps = 8/128 (6%)

35 Query: 1 MAYRKLGRSTSSQRKAMLRDLTIDLINESIVTTEARAKEIRKTVKRMITLQKRGDLHARR 60
 M+YRKLGRSTSSQRKAMLRDLTIDL+INE I TTE RAKS+R VERMITLQKRGDLHARR
 Sbjct: 1 MSTRKLGRSTSAQRKAMLRDLTIDLINERIEETTETRAKELRSVVERMITLQKRGDLHARR 60

Query: 61 QAAAYVRNEIASENYDEASDKYTSITLQKLFDDIAPRYAERNNGYTRILKTEPRRGDA 120
 QAAAY+RNE+AA+E ++ ALOKLF DIA RY ER GSYTRI+K PRRGD A
 40 Sbjct: 61 QAAAYIRNEVANKENQ-----DALQKLFSDIATRYERQGGYTRIMKLPORRGDA 112

Query: 121 PMAIIEIV 128
 PMAIIEIV
 45 Sbjct: 113 PMAIIEIV 120

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6999> which encodes the amino acid sequence <SEQ ID 7000>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1609 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/128 (97%), Positives = 127/128 (98%)

-2552-

Query: 1 MAYRKLGRITSSQKAMLRDLTTDLLINSEIVTTEARAKEIRKTVKRMITLGKRGDLHARR 60
 MAYRKLGRITSSQKAMLRDLTTDLLINSEIVTTEARAKEIRKTVKRMITLGKRGDLHARR
 5 Subjt: 1 MAYRKLGRITSSQKAMLRDLTTDLLINSEIVTTEARAKEIRKTVKRMITLGKRGDLHARR 60

Query: 61 QAAAYVRNELASENYDEASDKYTSTTALQKLEDDIAPRYAERNNGGYTRILKTEPRRGDAA 120
 QAAAYVRNELASENYDEA+DKYTSTTALQKLE +IAPRYAERNNGGYTRILKTEPRRGDAA
 Subjt: 61 QAAAYVRNELASENYDEADTKYTSTTALQKLPSEIAPRYAERNNGGYTRILKTEPRRGDAA 120

10 Query: 121 PMAIIEIV 128
 PMAIIEIV
 Subjt: 121 PMAIIEIV 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2265

A DNA sequence (GBSx2396) was identified in *S. agalactiae* <SEQ ID 7001> which encodes the amino acid sequence <SEQ ID 7002>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

20 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2384 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GF:AAA83977 GB:AF138877 mercuric reductase MerA [Bacillus sp.
 RC607]
 Identities = 29/33 (87%), Positives = 32/33 (96%)

Query: 4 VGLTEECAKEKGYDVKTSVLPLXAVPRAIVNR 36
 VGLTE+CAKEKGY+VKTSVLPL AVPRA+VNR
 35 Subjt: 520 VGLTECAKEKGYEVKTSVLPLDAVPRAIVNR 552

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2266

A DNA sequence (GBSx2397) was identified in *S. agalactiae* <SEQ ID 7003> which encodes the amino acid sequence <SEQ ID 7004>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

45 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3016 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]

-2553-

Identities = 146/194 (75%), Positives = 175/194 (89%)

Query: 2 PQISGLEKMDYLTSTTLLLELKKPKRLTVIGSGYIGMELQQLPHILQSEITLMQSRRL 61
 P I GL ++DYLTST+LLELKK+PKRL VIGSGYIGMELQQLPH+LQSE+TL+QSRRL
 5 Sbjct: 226 PNIPLGNEVDYLTSTTLLLELKKVFKRLVIGSGYIGMELQQLPHNLGSEVTLIQSRRL 285

Query: 62 KEYDPEISESVKALIEBQGINLVKGTATPERVQSGBIKKVFTVWGSREIVISDQLVAT 121
 KEYDPEISESVK+L+EQGINLVKGTAT+ER+EQ+G+IK+V+V VNG + +IE+DQLVAT
 10 Sbjct: 266 KEYDPEISESVKSLVBQGINLVKGTATYERIEBQNDIKKHEVEVNSKRIIEADQLVAT 345

Query: 122 GRKPNTDSLNLAAAGVETGKNNILINDPGQTSNEKIYAAGDVTLGFPQFVYVAAYEGGII 181
 GR PNT +LNL AAGVE G EI+I+D+ +T+N +IYAAGDVTLGFPQFVYVAAY+GG+
 15 Sbjct: 346 GRTPNTATLNLRAAGVETGSGEIIIDDYRTINIRIYAAGDVTLGFPQFVYVAAYGQVA 405

Query: 182 TUNAIGGLNKKIDL 195
 NAIIGLNKK++L
 15 Sbjct: 406 APNAIGGLNKKLNL 419

There is also homology to SEQ ID 1820.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2267

- A DNA sequence (GBSx2398) was identified in *S.agalactiae* <SEQ ID 7005> which encodes the amino acid sequence <SEQ ID 7006>. This protein is predicted to be triacylglycerol acylhydrolase. Analysis of this
 25 protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

- 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3180(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2268

- A DNA sequence (GBSx2399) was identified in *S.agalactiae* <SEQ ID 7007> which encodes the amino acid sequence <SEQ ID 7008>. Analysis of this protein sequence reveals the following:
 40

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0544(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 50 >GP:ARC74453 GB:AF000234 orf, hypothetical protein [Escherichia coli K12]
 Identities = 45/58 (77%), Positives = 51/58 (87%)

-2554-

Query: 1 MPWQNLHAGQENLPSGLTALTARPTVGBGKMTTDEPCSMAPDOKHDLISGTCSHLP 58
 +PWQNLHAG+ENLPSGLTALTARPTVGBG+IAM HD P APD+ DLISGTCSHLP
 Sbjct: 34 LPWQNLHAGEENLPSGLTALSAREPTTCEGELMAHDVPLGCAPEYDDLSGTCSHLP 91

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2269

- A DNA sequence (GBSx2400) was identified in *S.agalactiae* <SEQ ID 7009> which encodes the amino acid sequence <SEQ ID 7010>. This protein is predicted to be transposase for insertion sequence element is5. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2058 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB15497 GB:AK026530 unnamed protein product [Homo sapiens]
 Identities = 297/299 (99%), Positives = 297/299 (99%)

- Query: 1 MEQILFWQNMVEVIEFFYPKAGNGRPYPLETMLRIHCHQHWYNLSGDGAMEDALYEIASM 60
 MEQILFWQNMVEVIEFFYPKAGNGRPYPLETMLRIHCHQHWYNLSGDGAMEDALYEIASM
 Sbjct: 40 MEQILFWQNMVEVIEFFYPKAGNGRPYPLETMLRIHCHQHWYNLSGDGAMEDALYEIASM 99
- Query: 61 RLPARLSLDSALFDRITIMNFRHLLBQHQLARQLFKTINRWLAEGVMMTQSTILVDATTI 120
 RLPARLSLDSALFDRITIMNFRHLLBQHQLARQLFKTINRWLAEGVMMTQSTILVDATTI
 Sbjct: 100 RLPARLSLDSALFDRITIMNFRHLLBQHQLARQLFKTINRWLAEGVMMTQSTILVDATTI 159
- Query: 121 EAPSSTKNKEQQRDPFEMHQTCKKGNQHPGMAHIGVDKSGLTHSLVTTAANEHDLNQLK 180
 EAPSSTKNKEQQRDPFEMHQTCKKGNQHPGMAHIGVDKSGLTHSLVTTAANEHDLNQLK
 Sbjct: 160 EAPSSTKNKEQQRDPFEMHQTCKKGNQHPGMAHIGVDKSGLTHSLVTTAANEHDLNQLK 219
- Query: 181 NLLHGEQFVPSADAKYQGAPQREELASVDVDWGLAERPGKVRTLKQHPRNKNTAINIEYM 240
 NLLHGEQFVPSADA YQGAPQREELASVDVDWGLAERPGKVRTLKQHPRNKNTAINIEYM
 Sbjct: 220 NLLHGEQFVPSADAKYQGAPQREELASVDVDWGLAERPGKVRTLKQHPRNKNTAINIEYM 279
- Query: 241 KASIRARVEHPFRIIIRKQPGFVKARYKGLLKNENQLAMLPFLANLPADQMIRQWERSH 299
 KASIRARVEHPFRIIIRKQPGFVKARYKGLLKNENQLAMLPFLANLPADQMIRQWERSH
 Sbjct: 280 KASIRARVEHPFRIIIRKQPGFVKARYKGLLKNENQLAMLPFLANLPADQMIRQWERSH 338

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2270

A DNA sequence (GBSx2401) was identified in *S.agalactiae* <SEQ ID 7011> which encodes the amino acid sequence <SEQ ID 7012>. Analysis of this protein sequence reveals the following:

- Possible site: 16
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

-2555-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GF:CA851958 GB:AL109661 putative eukaryotic-type serine/threonine
 protein kinase [Streptomyces coelicolor A3(2)]
 Identities = 49/169 (28%), Positives = 90/169 (52%), Gaps = 6/169 (3%)

10 Query: 23 PTTIRVPDSNKTVAQAKMTLENSGLKVGAIKNIESDSVSGLVVKTDPAAGRSRREGAK 82
 P T+++EDV+ + QA+ LE+ GL+ G + SD V+ G V+ T P +G + R G+
 Sbjct: 469 PQTVKLPDVTGYKLDKARTLLEDEGLFEGMVTAPASDEVARGFVISTKPGSGTTVRAGSA 528

15 Query: 83 VNLVYIATPNKSPFTLGNKYKHNKYDKILKDLQKGVKSLIKVRKINNDYTTGTLAQSLP 142
 V L + + + + + L+G G+K + ++N+Y +G + A+ P
 Sbjct: 529 VAL-VVSGSPVDVDPVTDGDDLEARAELEGAGLK--VKTADERVNSYDSEGRV-ARQTP 584

20 Query: 143 EGTSFNPQGNKKLTLTVAVNDPMI-MPDVTGMTVGEVETLTDLGLAD 190
 E +G+ +TLTV+ MI +EDV G +V + + L D G + D
 Sbjct: 585 EPGGRAAAGD-TVTLTVSGKPMIEVDVVGDEVDAAKQKLEADAGFEVD 632
 Identities = 45/161 (27%), Positives = 80/161 (48%), Gaps = 4/161 (2%)

25 Query: 27 RVDPVSNKTVAQAKMTLENSGLKVGAIKNIESDSVSGLVVKTDPAAGRSRREGAKVNLV 86
 +VP + +KT AQA+ L+++GL VG +R+ SD+V G V+ TDP G R+ V+L
 Sbjct: 405 KVFPLLEKTRAAQAROLDADAGLDVGKVRFAYSDTVERGKVISTDPGVGRIRKNDSESLT 464

30 Query: 87 IATPNKSPFTLGNKYKHNKYDKILKDLQKGVKSLIKVRKINNDYTTGTLAQSLP 146
 ++ + L + + + L+ +G+ + V R +++ G +++ GT+
 Sbjct: 465 VSDGPTVKLPDVTGKLDKARTLLEDEGLFEGM--VTPASDEVARGFVISTKPGSGTT 522

Query: 147 FNPQGNKKLTLTVAVNDPMI-MPDVTGMTVGEVETLTDLGL 187
 + L V+ P+ +PDVTG + E L GL
 Sbjct: 523 VR--AGSAVALVSGSPVDVDPVTDGDDLEARAELEGAGL 561

35 There is also homology to SEQ ID 3026.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2271

40 A DNA sequence (GBSx2402) was identified in *S. agalactiae* <SEQ ID 7013> which encodes the amino acid sequence <SEQ ID 7014>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9311> which encodes amino acid sequence <SEQ ID 9312> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GF:AAB90561 GB:AB001058 glutamine ABC transporter, ATF-binding
 protein (glnQ) [Archaeoglobus fulgidus]
 Identities = 142/219 (64%), Positives = 178/219 (80%)

Query: 1 MDIHQGEVVVIGPSGSGKSTFLRTNLLVPTKGTVPESGIDITDKNDIFKMRKMG 60
 M + +GRVVVIGPSGSGKST LR +N LR PT G + +G+DIT+ K DI K+R+++G+
 Sbjct: 24 MKVEKGQVVVIGPSGSGKSTLLRCINLRBEPTSGKILLGDVITNSKIDINKVRORIGI 83

-2556-

Query: 61 VFQGFNLFPHMTVLENTLSPKTKGLSNLDAQTKAYELLEKVLKKEKANTYPASLSGGQ 120
 VFQGFNLFPHMTVLENTLSPKTKGLSNLDAQTKAYELLEKVLKKEKANTYPASLSGGQ 120
 Sbjct: 84 VFQGFNLFPHMTVLENTLSPKTKGLSNLDAQTKAYELLEKVLKKEKANTYPASLSGGQ 143

Query: 121 QQRATATAGLAMNPVLLFDEPTSAIDPEMVGVLTVMDLAKSMGTMVTVTHRMGFARE 180
 QQRATATAGLAMNPVLLFDEPTSAIDPEMVGVLTVMDLAKSMGTMVTVTHRMGFARE 180
 Sbjct: 144 QQRATATAGLAMNPVLLFDEPTSAIDPEMVGVLTVMDLAKSMGTMVTVTHRMGFARE 203

Query: 181 VADRVIPMDAGIIVEQGAPEVFEQTKKIRTRDFLSKVL 219
 VADRVIPMDAGIIVEQGAPEVFEQTKKIRTRDFLSKVL 219
 Sbjct: 204 VADRVIPMDAGIIVEQGAPEVFEQTKKIRTRDFLSKVL 242

There is also homology to SEQ ID 1186.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2272

A DNA sequence (GBSx2403) was identified in *S.agalactiae* <SEQ ID 7015> which encodes the amino acid sequence <SEQ ID 7016>. This protein is predicted to be 4-hydroxy-2-oxoglutarate aldolase (kdgA).

- 20 Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.1479 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 30 >GP:CAB14127 GB:299115 deoxyphosphogluconate aldolase [Bacillus subtilis]
 Identities = 21/62 (33%), Positives = 38/62 (60%), Gaps = 4/62 (6%)
- Query: 3 QLMQGRKIVAVIRGNSQEEAFQAAQACINGGISAIEIATYNSKASQVITQILVTOYTNQSGV 62
 +L + K + +AVIR ++EA Q ++ + GI A+E+ YT AS +IE + N+E +
 35 Sbjct: 9 RLKEARGLIAVIRSKDKQACQQTIESLCKGIRAVEVYTTTPGASDIIE---SPFRASDI 64
- Query: 63 VV 64
 ++
 40 Sbjct: 65 LI 66

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2273

A DNA sequence (GBSx2405) was identified in *S.agalactiae* <SEQ ID 7017> which encodes the amino acid sequence <SEQ ID 7018>. This protein is predicted to be H repeat-associated protein (rfbQRS) (b1458). Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.0207 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

There is homology to SEQ ID 504.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2274

- 5 A DNA sequence (GBSx2406) was identified in *S.agalactiae* <SEQ ID 7019> which encodes the amino acid sequence <SEQ ID 7020>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -6.74    Transmembrane    2 - 18 ( 1 - 21)
10    INTEGRAL    Likelihood = -3.03    Transmembrane    73 - 89 ( 73 - 92)

----- Final Results -----
          bacterial membrane --- Certainty=0.3697 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 3376.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2275

A DNA sequence (GBSx2407) was identified in *S.agalactiae* <SEQ ID 7021> which encodes the amino acid sequence <SEQ ID 7022>. This protein is predicted to be insertion element IS1 protein InsB (insB_5). Analysis of this protein sequence reveals the following:

```

25    Possible site: 52
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4280 (Affirmative) < succ>
30          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2276

A DNA sequence (GBSx2409) was identified in *S.agalactiae* <SEQ ID 7023> which encodes the amino acid sequence <SEQ ID 7024>. Analysis of this protein sequence reveals the following:

```

40    Possible site: 13
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3937 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2277

- 5 A DNA sequence (GBSx2410) was identified in *S.agalactiae* <SEQ ID 7025> which encodes the amino acid sequence <SEQ ID 7026>. This protein is predicted to be triosephosphate isomerase (tpi). Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence
10  INTEGRAL    Likelihood = -0.37    Transmembrane    35 - 51 ( 35 - 51)

----- Final Results -----
          bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC43268 GB:D07640 triosephosphate isomerase [Lactococcus
lactis]
20  Identities = 50/75 (66%), Positives = 61/75 (80%)

Query: 6  IAGNWMQKKNPEEAKAFIEAVASKLPSSSELVEAGIAAPALTSTVLEAAKQSELKIAACN 65
          IAGNWMQK  EA+AF+EAV + LPSS+ VE+ I APAL L+ + +GSELK+AR+N
Sbjct: 7  IAGNWMQKTLSEAAQAFVEAVKQLPSSSDNVBSVIGAPALFLAPWYLRQSELKLAEN 66

25  Query: 66 SYFENSQAFTGENSP 80
          SYFN+GAFTGENSP
Sbjct: 67 SYFNAGAFTGENSP 81

```

- 30 There is also homology to SEQ ID 6838:

```

          Identities = 58/77 (75%), Positives = 68/77 (87%)

Query: 6  IAGNWMQKKNPEEAKAFIEAVASKLPSSSELVEAGIAAPALTSTVLEAAKQSELKIAACN 65
          IAGNWMQKKNP-EAKAF+EAVASKLPs+LV+ +AFA+ L T +EAAK S LK+AAQN
35  Sbjct: 7  IAGNWMQKKNPQEAQAFVEAVASKLPSTLDVDAVAAPVDLVTTIEAKDSVLKVAQN 66

Query: 66 SYFENSQAFTGENSPKV 82
          YFN+GAFTGE SPKV
40  Sbjct: 67 CYFENIGAFTGETSPKV 83

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2278

- 45 A DNA sequence (GBSx2412) was identified in *S.agalactiae* <SEQ ID 7027> which encodes the amino acid sequence <SEQ ID 7028>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence
          INTEGRAL    Likelihood = -2.39    Transmembrane    96 - 112 ( 96 - 112)

50  ----- Final Results -----
          bacterial membrane --- Certainty=0.1956 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```


-2559-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAAL4368 GB:D50354 surface protein antigen precursor
[Streptococcus sobrinus]
Identities = 60/129 (46%), Positives = 76/129 (58%), Gaps = 18/129 (13%)
5
Query: 3 ISFDNSFLFETVSDGAPQADVYLQMKRIAAGQVNTYLYLTVNGYVSSNTVVTHTPQPRE 62
+F FL +VS DSAQCA+VYLQMKRIA G NTY+TVNG SSNTV T TP+P++
Sbjct: 1442 VTFKEDFLRSVSDGAPQAEVYLQMKRIAGTGFANTYVNTVNGITYSSNTVTRTSTPEPQ 1501
10
Query: 63 PSPNQP-----TPPQPIETIEPPVPSILHNTGQES---LGLIG--AGILLGT 108
PSP P P Q PP A LD TG+ + LGL+ AG L
Sbjct: 1502 PSPVDKTKTTTVFPQPRCKFAYQPPAPGAG--LPATGSSNAYLPLGLVSLTAGFS-- 1558
15
Query: 109 AYGLKKKEE 117
GL+K++
Sbjct: 1559 -LGLRRKQD 1566

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
20 vaccines or diagnostics.

Example 2279

A DNA sequence (GBSx2413) was identified in *S.agalactiae* <SEQ ID 7029> which encodes the amino acid sequence <SEQ ID 7030>. Analysis of this protein sequence reveals the following:

Possible site: 23
25 >>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.3691(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9359> which encodes amino acid sequence <SEQ ID 9360> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAR15793 GB:Z99123 phosphotransacetylase [Bacillus subtilis]
Identities = 131/221 (59%), Positives = 169/221 (76%), Gaps = 2/221 (0%)
35
Query: 6 LVDPIVLKGADEVDHSLARIGFVDQVYSIIDPQYKFEEMGEAFVIRKKGATMEDADR 65
+++P++G +E+ L I DP YE E++ +AFVE RKGKAT E A +
40 Sbjct: 41 VLNPVIVIGNENSIQAKAKELNLTIGGVKIYDPHYEGMEDLVQAFVRRKKGKATEBQARK 100
Query: 66 LAKDVYFQVGLVGLGLADGVSGAHSHTADTVRPAQLIHTKPGISPTSCVFLMRENT 125
L D NYFG MLN GLADG+VSGA HSTADTVRPAQLIHTK G+ +TSQVP+H R
50 Sbjct: 101 ALLDENYFGTLMVYNSGLADGLVSGAHSHTADTVRPAQLIHTKRGVKTSGVFMARG-- 158
Query: 126 CERYIFADCAINIDENAGELAEIAVNTADTAKIFDIDPKIAMLSPSTKSKAKAFQAEKVQ 185
+E+Y+ADCAINI P++Q+LAELIA+ +A+TAK+FDI+P++MLSPSTKSAK+ +E+V
Sbjct: 159 ESQYIFADCAINIAPDSQDLAEIATASANTAKMFDIEEAVMLSPSTKSKAKSDETEKVA 218
60
Query: 186 EAAKIAKDLSPGLAVDGEIQLQFDAAPFVETARIKAPNSDVAG 226
+A KIAK+ +EEL +DKE QVDAAPV RE KAP+S++ G
Sbjct: 219 DAVKIAKEKAPFLIDGEPQFDAAPFVSEAEKAPDSEIKG 259

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7031> which encodes the amino acid
55 sequence <SEQ ID 7032>. Analysis of this protein sequence reveals the following:

Possible site: 34

-2560-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3182 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 181/227 (79%), Positives = 211/227 (92%)

10 Query: 1 MKFEGELVDPVILGKADEVHDSLARLGFVLDQYSDIDPQYKSEVHMEKAPVSRKHKATM 60
 +KFEGL+P+ILG++EV + L +IGF DQDY+II+P +Y P+MKRAFVE+RKKGAT+
 Sbjct: 38 LKFEGLLEPILGQSEVRNLLTKLGADQDTYITINNEYADFDMEKAPVSRKHKATL 97

15 Query: 61 EDADRLKRDVNYFGVMLVKLGADGMVSGAIHSTADTVREALQIIKTKPGISRTSGVFLM 120
 EDAD++L+DNNYFGVMLVK+GLADGMVSGAIHSTADTVREALQIIKTKPGISRTSGVFLM
 Sbjct: 98 EDADKMLRDVNYFGVMLVKLGADGMVSGAIHSTADTVREALQIIKTKPGISRTSGVFLM 157

20 Query: 121 NRENTCERYIFADCAINIDPFAQELAEIAVNTATAKIFDIDPKIAMLSPTKSGKAPQ 180
 NRENT ERY+PADCAINIDP AQELAEIAVNTA+TAKIFDIDPKIAMLSPTKSG KAPQ
 Sbjct: 158 NRENTSERYVFPADCAINIDPFAQELAEIAVNTAETAKIFDIDPKIAMLSPTKSGKAPQ 217

Query: 181 AEKQCEPAKIAKDLSPELAVDGELOQDPAFVETARIKAPNSVAGK 227
 +KVAEA +IA L+P+LA+DGELOQDPAFVETARIKAPNSVAG+
 25 Sbjct: 218 VDKVREATEIATGLNPLDALDGELOQDPAFVETARIKAPNSVAGQ 264

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2280

30 A DNA sequence (GBSx2414) was identified in *S. agalactiae* <SEQ ID 7033> which encodes the amino acid sequence <SEQ ID 7034>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.4076 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AG19110 GB:AB005009 Vnc0600c [Halobacterium sp. NRC-1]
 Identities = 57/176 (32%), Positives = 86/176 (48%), Gaps = 20/176 (11%)

45 Query: 1 MKVLLYLEAEVLYKSGSIGRAIKHQKALQIAGIDYTTNPT----- 41
 M+ L YLEA E L+ G+ A Q AL+ ++ P
 Sbjct: 2 MRALNYLEAEALR-GGMVATNQRAALEITDVEVETVNRAGDPVRSIGSLAAGGSCF 60

50 Query: 42 DDFDLVNNITYGIRSKLLMSKAKIKKKVIMHGHSREDFNSFIQSNLVSPLFQNYLCR 101
 FD+ R N G S + A+T+ ++H H T EDF SF GS+ ++P + YL
 Sbjct: 61 TAPDVAKNDLVGKSVAVARHARTITPLVLAHATREDFAQSPGSGSTADAPALPYLRW 120

Query: 102 FYQKADAIITFDYKQLIKAYGIKKPIFVLSNMDLSRYCKSEKKESAPHYVHL 157
 FY +AD ++P++Y+K +++AY + PI LSNQ+DL Q R + R F L
 55 Sbjct: 121 FYSQADLVLCPSBYTKDVLRAYPVDAPIRQLSNQVDLESMQGVESFRADTRAPDL 176

There is also homology to SEQ ID 1220.

-2561-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2281

A DNA sequence (GBSx2415) was identified in *S.agalactiae* <SEQ ID 7035> which encodes the amino acid sequence <SEQ ID 7036>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2625 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC35010 GB:AF055987 intracellular a-amylase [Streptococcus mutans]
Identities = 27/46 (58%), Positives = 33/46 (71%)

Query: 1 MEVGEIYAGRTFPVDYLGNCBQEVVIGDDGNCDFLVESASISAMVVK 46
      M +GE K FVDYL NC +EV++ D GNCDP V+ AS+SAMV K
Sbjct: 438 MNMGEPFRNKVFVDYLNKCTREVILODQGNCFPFVQASLGAMVVK 483
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2282

A DNA sequence (GBSx2416) was identified in *S.agalactiae* <SEQ ID 7037> which encodes the amino acid sequence <SEQ ID 7038>. This protein is predicted to be RopA. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2082 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

There is also homology to SEQ ID 6908:

```
Identities = 30/35 (85%), Positives = 33/35 (93%)

Query: 1 MEADQVRGLLSADMLKHDIAMKAVDVITTSATVK 35
      M ADQVR LLSADMLKHDIAMKAV+VITTS+AA+VK
Sbjct: 422 MPADQVRSLLSADMLKHDIAMKAVKVVITTSATVK 456
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2283

A DNA sequence (GBSx2417) was identified in *S.agalactiae* <SEQ ID 7039> which encodes the amino acid sequence <SEQ ID 7040>. This protein is predicted to be DNA-directed RNA polymerase, subunit delta. Analysis of this protein sequence reveals the following:

-2562-

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2407 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CA15744 GB:Z99123 RNA polymerase (delta subunit) [Bacillus subtilis]
 Identities = 62/186 (33%), Positives = 102/186 (54%), Gaps = 15/186 (8%)
 Query: 1 MELEVFAGQKSELSMIEVARAILLEQRGRDNEMYPFSLVNDIQTLYGKSDSAIRESLFF 60
 M ++ ++ +E E++++E+A + E+ + + F +L+N+I + LG + + + F
 15 Sbjct: 1 MGIRQYSQERLKEMLVSEIAHELFEHKKP--VPFQELNLSIASLLGVKKEELGDIRIAQF 58
 Query: 61 YSDLNTDGSFIP LGENKGLRSWYAIDRIDSKITILEEDEDGAPKRRKKRVNAPMGQDED 120
 Y+DLN DG F+ L + WGLRWSY D+DEE K KKK+ ++ D D
 20 Sbjct: 59 YTDLNTDGRFLALSDQITWGLRWSYFDQLDEE-----TQPTVKAKKKKAKKAVESOLD 111
 Query: 121 AIDYNDODPEDEDFTETPSLEYDERNPDDERKSEVSEYSDSINEIIPDEDLDEUVEINEE 180
 ++ + D +D D E L+ + + ++D E + + D E I E I DED DED
 25 Sbjct: 112 LDFPERIDEDDIDLDEVESEKLDLRADDPDEEDLORDODDLRETDIEDD-DEUY----- 165
 Query: 181 DDEEEE 186
 DDEEEE
 30 Sbjct: 166 DDEEEE 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7041> which encodes the amino acid
 30 sequence <SEQ ID 7042>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2263 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

40 Identities = 162/191 (84%), Positives = 181/191 (93%), Gaps = 1/191 (0%)
 Query: 1 MELEVFAGQKSELSMIEVARAILLEQRGRDNEMYPFSLVNDIQTLYGKSDSAIRESLFF 60
 ++L+VFAGQKSELSMIEVARAILLE+RGRDNEMYPFSLVNDI+Q YLGSKD+ IR +LFF
 45 Sbjct: 12 LKLDVFAGQKSELSMIEVARAILLEQRGRDNEMYPFSLVNDI+QNYLGKSDAGIRHRLFF 71
 Query: 61 YSDLNTDGSFIP LGENKGLRSWYAIDRIDSKITILEEDEDGAPKRRKKRVNAPMGQDED 120
 Y+DLNTDGSFIP LGENKGLRSWYAIDRIDSKITILEEDEDGA KRRKKRVNAPMGQDED
 50 Sbjct: 72 YTDLNTDGSFIP LGENKGLRSWYAIDRIDSKITILEEDEDGAPKRRKKRVNAPMGQDED 131
 Query: 121 AIDYNDODPEDEDFTETPSLEYDERNPDDERKSEVSEYSDSINEIIPDEDLDEUVEINEE 180
 AIDY DDDPEDEDFTET+ +YDEH+PDDERKSEVSEYSDS+NEIIP++D E+V+INEE
 Sbjct: 132 AIDYRDDDPEDDDFTETSAEVEYDREDPDDERKSEVSEYSDSLNEIIPDDDF-EEVDINEE 190
 Query: 181 DDEEEEEEEV 191
 D+E+EE+EE V
 55 Sbjct: 191 DERDDEDEEFV 201

Based on this analysis, it was predicted that these proteins could be useful antigens for
 vaccines or diagnostics.

Example 2284

A DNA sequence (GBSx2418) was identified in *S. agalactiae* <SEQ ID 7043> which encodes the amino acid sequence <SEQ ID 7044>. This protein is predicted to be CTP synthetase (pyrG). Analysis of this protein sequence reveals the following:

```

5      Possible site: 23
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -0.11    Transmembrane    5 - 21 ( 5 - 21)

10      ----- Final Results -----
      bacterial membrane --- Certainty=0.1044 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:CAA09021 GB:AJ010153 CTP synthetase [Lactococcus lactis subsp.
      cremoris] (ver 2)
      Identities = 421/533 (78%), Positives = 481/533 (90%) .

20      Query: 2   TKYIFVTGGVSSIGKGI VAASLGRLLKNRGLKVTIQKDFPINIDPGTMSPYQHGEVYV 61
      TKYIFVTGG SS+GKGI VAASLGRLLKNRGLKVT+QKDFPINIDPGTMSPYQHGEV+V
      Sbjct: 3   TKYIFVTGGTSSMGKGI VAASLGRLLKNRGLKVTQKDFPINIDPGTMSPYQHGEVVF 62

30      Query: 62   TDGAGTDLDLGHYERFIDINLNKYSNVTGKIYSEVLKKBRRGELGATVQVIPHVDA 121
      TDGAGTDLDLGHYERFIDINLNKYSNVT+GK+YSE+L+KER+GEYL GATVQ++PHVT+
      Sbjct: 63   TDGAGTDLDLGHYERFIDINLNKYSNVTSGKVYSELRLKRRKGEYL GATVQVMPHTNN 122

40      Query: 122  LKEKIKRAATTTSDVITTEVGTVGDIESLPFLEALRQMKADVGSNDNMYIHTILLPYL 181
      LKEKIKRAATTTD+D+IITEVGTVGD+ESLFF+EALRQMK+VQ+DNVMIYHT + +L
      Sbjct: 123  LKEKIKRAATTTDADIIITEVGTVGDIESLPFLEALRQMKAEVGDNMYIHTVPIHL 182

50      Query: 182  KAAGEMTKTKPTQHSVKELRGLGIQPMILVIRTEQPAQGSINNLQAQPCDVAPEAVIESLD 241
      +AAGE+KTK Q++K LR GIQ NMILV+R+E P +++K+A PCDV APEAVI+SLD
      Sbjct: 183  RAAGELTKTKIAQNTKTLREYGIQANMLVLRSEVPTTMRDKIAMPQDVAPEAVIQVFE 242

60      Query: 242  VDHIYQIPLNMQAQNMDQIVCDHLKLETPAADMTENSAMVDKVMLEKKVKIALVGKYVE 301
      V+H+YQIPLN+QAQNMDQIVCDHLK+ P ADM EWSAMVD VMEL+KVKIALVGKYVE
      Sbjct: 243  VEHLIYQIPLNQAQNMDQIVCDHLKLDAPKADMAEWSAMVDHVMLEKKVKIALVGKYVE 302

70      Query: 302  LPDAYLSVVEALKHSGYVNDVAIDLKGVNAAEVTEDNIELVGDAGIIVPGGFGQRGE 361
      LPDAY+SV RALKH+GY +D +D+ MVA +VT++N+ ELVGDAGIIVPGGFGQRG+E
      Sbjct: 303  LPDAYISVTEALKHAGYASDAEVDINNVANDVTDENVAELVGDAGIIVPGGFGQRGE 362

80      Query: 362  GKIEAIRYARENDVPMGLGVLGMQLTVCVEFARNVNLHGAASAEIDPKTFPIIDIMRDQ 421
      GKI AI+YARENDVPMGL+CLGMQLT VEFARNVL L GA+S ELDP+T +P+IDIMRDQ
      Sbjct: 363  GKIAAIKYARENDVPMGLGICLGMQLTVAEFARNVNLGEGHSEHSEFDPETKYVIDIMRDQ 422

90      Query: 422  IDIEDMGTTLRGLYPCKLKSGRSAAAYNNQEVVQRRHRHRYEFPNTKFRQGEAAGVF 481
      +D+ELMGTTLRGLYP KLK+GSRA AAYN+ EVVQRRHRHRYEFPN K+RE FE AGVF
      Sbjct: 423  VDVEDMGTTLRGLYPACKLKSGRSAAAYNDAEVVQRRHRHRYEFPNKKYREDFEKAGVF 482

100     Query: 482  SGVSPDNRLMEVVELPEKKFFVAQYHFELQSRPNHAEELYTAFTYAAVENMK 534
      SGVSPDNRL+E+VRL KKKFFVA QYHFELQSRPN ERLYT F+ AVEN K
      Sbjct: 483  SGVSPDNRLVEIVELSGKKFFVAQYHFELQSRPNRPEELYTEFIRVAENSK 535

```

55 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 7045> which encodes the amino acid sequence <SEQ ID 7046>. Analysis of this protein sequence reveals the following:

```

      Possible site: 23
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -0.11    Transmembrane    5 - 21 ( 5 - 21)

60      ----- Final Results -----
      bacterial membrane --- Certainty=0.1044 (Affirmative) < succ>

```

-2564-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:CAA09021 GB:A010153 CTP synthetase [Lactococcus lactis subsp.
cremoris] (ver 2)
Identities = 423/532 (79%), Positives = 483/532 (90%)

10 Query: 2 TKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDOPYINIDPGTMSPYQHGEVYV 61
TKYIFVTGG SS+KGKIVAASLGRLLKNRGLKVTIQKFDOPYINIDPGTMSPYQHGEV+V
Sbjct: 3 TKYIFVTGGTSSMGKIVAASLGRLLKNRGLKVTIQKFDOPYINIDPGTMSPYQHGEVVF 62

Query: 62 TDDGAETDLDLGHYERFDININIKYSNVTGKIYSEVLKRRKRGYLGATVQVIPHITDA 121
TDDGAETDLDLGHYERFDININIKYSNVT+GK+YSE+LRRKRRGYLGATVQ++PH+T+
15 Sbjct: 63 TDDGAETDLDLGHYERFDININIKYSNVTSGKIYSEILRRKRGYLGATVQMPVPHVTNN 122

Query: 122 LKEKIKRAASTTDSVLIITEVGOTVGDIESLFFLEALRQMKADVGSERNVYIHTILLPYL 181
LKEKIKRAA+TTD+D+IITEVGOTVGD+ESLFF+EALRQMK+VG++NVYIHT +L
Sbjct: 123 LKEKIKRAAITTDADIIITEVGOTVGDMSLFFLEALRQMKAEVDGNVYIHTVPIHL 182

20 Query: 182 KAAGEMTKPTQHSVKELRGLGIQPNMLVIRTEEPVQGIKNKLAQPCDVNSEAVIESRD 241
+AAGE+KTK Q++K LR GIQ NMLV+R+E P+ +++K+A PCDV EAVI+S D
Sbjct: 183 RAAGELTKTIAQNTATKLTREYGIQANMLVLRSEVPITTEMRDKIANPCDVAPEAVIGSLD 242

25 Query: 242 VEHLQYIPLNLQAQNMQIVCDHLKLAPADMTEWSAMVDKVMNLKRTTKIALVGKGYE 301
VEHLQYIPLNLQAQ+MDQIVCDHLKL+AP+ADM EWSAMVD VMNL+K KIALVGKGYE
Sbjct: 243 VEHLQYIPLNLQAQNMQIVCDHLKLDAPKADMEWSAMVDHVMNLKKKVKIALVGKGYE 302

Query: 302 LPDAYLSVVAELKHSGYANDTAIDLQVNVANDVTVDNAADLLGDAGIIVPGGGGQRGTE 361
LPDAY+SV EALKH+GYA+D +D+ NVNANDVT +N A+L+GDA GIIVPGGGGQRGTE
Sbjct: 303 LPDAYISVTEALKHAGYASDAEVDINNVNANDVTDENVAELVGDAIGIIVPGGGGQRGTE 362

35 Query: 362 GKIAIRYARENDVPMGLIGLGMQLTVEFARHVLNMBGANSFELPSTKYPIIDIMRDQ 421
GKI AI+YARENDVPMGLIGLGMQLT VEFAR+VL +EGA+SEFL+P TKYP+IDIMRDQ
Sbjct: 363 GKIAIRYARENDVPMGLIGLGMQLTAVEFARNVLGLEGAHS FELDPSTKYPIIDIMRDQ 422

Query: 422 IDIEDMGGTLRLGLYPCKLKPGSKAAMAYNNQEVQRRRHRHRYEPNNKRFPEFAAGFVF 481
+D+EDMGGTLRLGLYP KLG GS+A AYN+ EVQRRHRRHRYEPNNK+R +FE AGFVF
40 Sbjct: 423 VDVEDMGGTLRLGLYPAKLNGSKAAYNDAEVQRRRHRHRYEPNNKREDPEKAGFVF 482

Query: 482 SGVSPDNLVSEIVELKEKKFFVAQYHPELQSRPNRPEELYTAFVTAIKNK 533
SGVSPDNLVSEIVEL KKFVA QHPELQSRPNRPEELYT P+ A++NS
Sbjct: 483 SGVSPDNLVSEIVELSGKKFFVAQYHPELQSRPNRPEELYTETFRIVAVNS 534

45

An alignment of the GAS and GBS proteins is shown below.

Identities = 477/532 (89%), Positives = 503/532 (93%)

50 Query: 1 MTKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDOPYINIDPGTMSPYQHGEVY 60
MTKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDOPYINIDPGTMSPYQHGEVY
Sbjct: 1 MTKYIFVTGGVSSIGKGIVAASLGRLLKNRGLKVTIQKFDOPYINIDPGTMSPYQHGEVY 60

Query: 61 VTDDGAETDLDLGHYERFDININIKYSNVTGKIYSEVLKRRRGEYLGATVQVIPHVD 120
VTDDGAETDLDLGHYERFDININIKYSNVTGKIYSEVL+KER+GEYLGATVQVIPH+TD
55 Sbjct: 61 VTDDGAETDLDLGHYERFDININIKYSNVTGKIYSEVLKRRRGEYLGATVQVIPHVD 120

Query: 121 ALKEKIKRAASTTDSVLIITEVGOTVGDIESLFFLEALRQMKADVGSERNVYIHTILLPY 180
ALKEKIKRAA+TTDSDVLIITEVGOTVGDIESLFFLEALRQMKADVGS+ERNVYIHTILLPY
Sbjct: 121 ALKEKIKRAASTTDSVLIITEVGOTVGDIESLFFLEALRQMKADVGSERNVYIHTILLPY 180

60 Query: 181 LKAAGEMTKPTQHSVKELRGLGIQPNMLVIRTEQPAQGSIKNKLAQPCDVAPAEVIESL 240
LKAAGEMTKPTQHSVKELRGLGIQPNMLVIRTE+P Q IKNKLAQPCDV EAVIES
Sbjct: 181 LKAAGEMTKPTQHSVKELRGLGIQPNMLVIRTEEPVQGSIKNKLAQPCDVNSRAVIESR 240

65 Query: 241 DVDHIYQIPLNLQAQNMQIVCDHLKLSTPAADMTEWSAMVDKVMNLKRLKIALVGKGY 300
DV+H+YQIPLNLQAQ+MDQIVCDHLKL P ADMTEWSAMVDKVMNL KIALVGKGY

-2565-

Sbjct: 241 DVSEHLYQIPLMLQAQSQDQIVCDHIALNAPQADMTESAMVDKVMNLKIKTKIALVGKYY 300
 Query: 301 ELPDAYLSVVEALKHSGYVNDVAIDLKQVNAABVTENIKELVQDADGIIIVPGGFGQGRG 360
 ELPDAYLSVVEALKHSGY ND AIDLKQVNA +VT DN +L+QDADGIIIVPGGFGQGRG+
 Sbjct: 301 ELPDAYLSVVEALKHSGYANDTAIDLKQVNAANDVTDMAADLLQDADGIIIVPGGFGQGRG 360
 Query: 361 EGKIEAIRYARENDVPMGLGVCLGMQLTVEFARNVLMHGANSAELDPKTPFPPIIDIMRD 420
 EGKI+AIRYARENDVPMGLG+CLGMQLTVEFAR+VLM+ GANS EL+P T +PIIDIMRD
 Sbjct: 361 EGKIQAIRYARENDVPMGLGICLGMQLTVEFARHVLAMEGANSFELEPSTKYPIIDIMRD 420
 Query: 421 QIDIDEMGTTLR/LGLYPCKLKSGSRAAAYNNQEVVQRHHHRYEPNTKFRQGFAAGFV 480
 QIDIDEMGTTLR/LGLYPCKLK GS+AA AYNQEVVQRHHHRYEPN KFR +FEAGFV
 Sbjct: 421 QIDIDEMGTTLR/LGLYPCKLKSGSKAAMAYNNQEVVQRHHHRYEPNNKFRPEFEAGFV 480
 Query: 481 FSGVSPDNRLMEVVKLPKPKFFVAAQYHPELQSRPNHAEKLITAFVTAIVEN 532
 FSGVSPDNRL+E+VEL EKKFFVAAQYHPELQSRPN EELITAFVTA++N
 Sbjct: 481 FSGVSPDNRLVIVELKEKKFFVAAQYHPELQSRPNRPELELITAFVTAIAKN 532

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2285

A DNA sequence (GBSx2419) was identified in *S. agalactiae* <SEQ ID 7047> which encodes the amino acid sequence <SEQ ID 7048>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.92 Transmembrane 13 - 29 (3 - 34)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9285> which encodes amino acid sequence <SEQ ID 9286> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14296 GB:299116 ygd [Bacillus subtilis]
 Identities = 79/289 (27%), Positives = 139/289 (47%), Gaps = 8/289 (2%)
 Query: 1 MKKIRLSKFKIMIVVILFLISVAASFFVHVAQVRDDKSFISNGRKPNGLYAYDKSFD 60
 MKKI L+ I +V + I + S ++ + D+ I + G+ ++ +SF+
 Sbjct: 1 MKKILLA--IGALMTAVIAIGVIFSMILFLIKKTD--IIKRETNHGVF--ESFE 53
 Query: 61 KLLKQKIHTNQIKQVAWTVPAVKTHKTAVVVHGFANSKEMHAYGNLPHKLGYNVLM 120
 ++ K + + + Y A T T ++ HG + N Y LF LG+NVL+
 Sbjct: 54 QMKTAFVIPAAGYDINGKTHVAPHTNTIIICGVTMNVLSIKYMHFLDLGVRVLI 113
 Query: 121 PDNIHGESHQQLIGYGMNDRENIIKWTMIVDK-NPSSQITLFGVSHGAGVVMASGEK 179
 D+ HG+S G+ YG+ ++++ K ++ +K N I + G SHG T ++ +G
 Sbjct: 114 YDHRHGGSGGKTSYGFPYKDDINKVVSLLKNKYNHRRGLIGIHGESHGAVTALLYAGAH 173
 Query: 180 LPSQVMMIIEDCGYSSVDELKPKQKEMTGLFAPFLLYEVETISKIKAGFSYGQASSVEQ 239
 I DC ++ ++L ++ + Y LP+PLL K+R G+ + S +
 Sbjct: 174 CSDGADFYTDACPFACFDEQLAYRLAETLYLPSWFLPIADFFLKLGGYRAREVSPVAV 233
 Query: 240 LKNNLDPALPIHGDKDNFVPTSMVYNYKATAGKKELYIVKGAKIAPSS 288
 + K P LFIH D++P S Y+ G K LYI + +HA S+
 Sbjct: 234 IDKIEKPVLFISKDDYIPVSSSTERYEKRGKPKALYIAENGHEAMSY 282

-2566-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7049> which encodes the amino acid sequence <SEQ ID 7050>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.48 Transmembrane 10 - 26 (3 - 32)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3994 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP_CAB14296 GB:299116 yqkD [Bacillus subtilis]
 Identities = 88/295 (29%), Positives = 145/295 (48%), Gaps = 4/295 (1%)
 Query: 10 LGILFLITLISVGASFYFPHVAQIREEKSPINNKRSTNNPLYPABQSDALPYEKRL 69
 L I L+ +I+G F H+ ++K+ + KR T-N + +SF+ + +
 Sbjct: 6 LAIGALVIAVIAIG--IVFSHMILFIKKTDIEDIIRKTDNG-HDVPESFEQMEKTAFTVI 62
 Query: 70 TNRGLQVGVNLPAAQKTKYKTAIVVHGFTNDKEDMKPYAMLFDLGYNVLMFDNEAHGES 129
 + + Y A T T I+ HG T + + Y LF DLG+NVL+ D+ HG+S
 Sbjct: 63 PSAYGYDIEGYVAPHDPTNTIIICGVTMVNLNGLKYNHLDLGNVLIYDHRHQS 122
 Query: 130 EGNLIGYGNDRINLVNVAWTDLI-KENPESQITLFGLSMGAAATVMMASGERLPAQVTSILI 188
 G YG+ ++ ++ L K N I + G SMGA T ++ +G
 Sbjct: 123 GGRKTSYGYFEKDLKLVSVLLNKNTHRGRLIGIHGSGAVTALLYAGAHCSGDGAFYI 182
 Query: 189 EDCGYASVNDLKFQAKAMYNLPAFPLLYEVSALSIRAGFSYGEASSVQLAANGRPL 248
 DC +A ++L ++ +A Y LF+PLL K+R G+ ES + + K +P L
 Sbjct: 183 ADCPFACFDEQLAYRLAEYRLFSPMLLPFIADPFLLKRGYRAREVSPLAVIDRIEKFPVL 242
 Query: 249 FIHGDKDDFVPTMVDNYKATGKPKRILIVGAKHAKSFETNPQYQKKIAAPL 303
 FIH DD+P Y+ +GPK + I + +HA S+ N Y+K + FL
 Sbjct: 243 FIHSKDDYIPVSSTERLYEKGKPKALYLAENGHAMSTYKNRHYRTKTVQEP 297

An alignment of the GAS and GBS proteins is shown below.

Identities = 203/294 (69%), Positives = 246/294 (83%)
 Query: 1 MKKIRLSKFIKIMIVILFLISVARSFYFPHVAQVRDDKSPISNGORKPGNSLYAYDKSFD 60
 MK IR+K+ ++ +++ LISV ASFYFPHVAQ+R++KSF+N +R N LY ++SFD
 Sbjct: 1 MKTIRIAKYLGLILFLITLISVGASFYFPHVAQIREEKSPINNKRSTNNPLYPABQSD 60
 Query: 61 KLLKQKIEITNQNIKQVAVYVPAVKTHKTAIVVHGFANSKEMKAYGWLPHKLGYNVLM 120
 L +K ++TN+ +EQV WY+PA +KT KTA+VVEGF N KE+MK Y LPH LGYNVLM
 Sbjct: 61 ALPYEKRLTNRGLQVGVNLPAAQKTKYKTAIVVHGFTNDKEDMKPYAMLFDELGYNVLM 120
 Query: 121 PDNIHGESHGQLIGYGNDRINLIKWTMIVDKNPSSQITLFGVSMGAAATVMMASGERL 180
 PDN AHGES G LIGYGNDR N++ WT+ ++ +NP SQITLFG+SMG ATVMMASGE+L
 Sbjct: 121 PDNRAHGESGNLIGYGNDRINLVNVAWTDLIKENPESQITLFGLSMGAAATVMMASGERL 180
 Query: 181 PSQVNVLLIEDCGYSSVNDLKFQAKAMYNLPAFPLLYEVSITISKIRAGFSYGEASSVQL 240
 P+QV ++IEDCGY+SVNDLKFQK MY LPAFPLLYEVS +SKIRAGFSYG+ASSV+QL
 Sbjct: 181 PAQVTSILIEDCGYASVNDLKFQAKAMYNLPAFPLLYEVSALSIRAGFSYGEASSVQL 240
 Query: 241 KNNLPALFIHGDKDNFVPTMVDNYKATGKPKRILIVGAKHAKSFETPEK 294
 KN P LFIHGDKD+VPT MYVDNYKAT G KE+ IVGAKHAKSFET PR+
 Sbjct: 241 AKNRKPTLFIHGDKDDFVPTMVDNYKATGKPKRILIVGAKHAKSFETNPQ 294

SEQ ID 9286 (GBS662) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 8-10; MW 63kDa) and in Figure 187 (lane 4; MW 63kDa).

GBS662-GST was purified as shown in Figure 237, lane 7.

-2567-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2286

A DNA sequence (GBSx2420) was identified in *S. agalactiae* <SEQ ID 7051> which encodes the amino acid sequence <SEQ ID 7052>. This protein is predicted to be aspartate--ammonia ligase (asnA). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2898 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9309> which encodes amino acid sequence <SEQ ID 9310> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
Identities = 246/300 (82%), Positives = 268/300 (89%)
```

```
Query: 1 MIDKLSIVEVQGPILSQVGDGMQDNLGIEHPVSVKVNLINPEAEFEVHSLAKMKRHTLA 60
+I++L I+EVQGPILSQVGHGMQDNLGIE V V V IP A FEVHSLAKMKRHTLA
Sbjct: 23 LIEQLGILIEVQGPILSQVGDGMQDNLGIEKAVQVNVKCIPINAVFEVHSLAKMKRHTLA 82
```

```
Query: 61 RFGNSEGGLFVHMKALRPDSDSLDPTHSVYVDQMWKEKVIPDGRRLDYLKETVKEIKY 120
RF F E BGLFVHMKALRPDSDSLDPTHSVYVDQMWKEKVIP+GRRL YLKETV IY+
Sbjct: 83 RNFKRGDEGLFVHMKALRPDSDSLDPTHSVYVDQMWKEKVIPDGRRLNFAYLKETVNSIYR 142
```

```
Query: 121 AIRLTLAVEARFDI SILPKRITTFIHTBELVEKVPDLSPKERENAIKEYGAVFLIGIG 180
AIRLTLAVEARFDI SILPK+ITP+H+E-LV++YFDLS KERENAI KEYGAVFLIGIG
Sbjct: 143 AIRLTLAVEARFDI SILPKQITPVHSEDLVKRYPDLSKKERENAIKEYGAVFLIGIG 202
```

```
Query: 181 GELADGKPHIDGRAPDYDDWTTSENGFKGLNGDILVQNEQLGTAFELSSMGIRVDEALK 240
G+L+DGKPHIDGRAPDYDDWTT SENG+KGLNGD ILVQV+QLG AFELSSMGIRVDE AL+
Sbjct: 203 GKLSDGKPHIDGRAPDYDDWTTSENGYKGLNGDILVQNEQLGTAFELSSMGIRVDESLAR 262
```

```
Query: 241 RQVVLGTGDEDRLEFEWHKTLGAGFFPLTIGGGIGQSRILANLARKHIGEVQSSVWPKSV 300
QV LTGDED L+ +WH+ LL G PLTIGGGIGQSRILAN LARK HIGEVQSSVWPK+
Sbjct: 263 LQVVLGTGDEDRHLDHWHQDLNGLKPLTITGGIGQSRILANMLLARKHIGEVQSSVWPKSM 322
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 7053> which encodes the amino acid sequence <SEQ ID 7054>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 189 - 205 ( 189 - 205)

----- Final Results -----
      bacterial membrane --- Certainty=0.1065 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
Identities = 255/330 (77%), Positives = 289/330 (87%)
```

```
Query: 1 MKKSFTHQEEISPVNITFTQYLIKALDVEVQGPILSRVGDGMQDNLGTEPNVSVNL 60
MKK+FI QQ+EISPVNITFTQ LI +L ++EVQGPILS+VG+GMQDNLG E V VNV
```

-2568-

Sbjct: 1 MKKTFILQQEISFVNTFTQMLIBQLGIIEVQGPILSQVNGMQDNLSGIEKAVQVNVK 60

Query: 61 KIPNATFEVHSLAKWKRHTLARFGVNEGBGLVNMKALRPDESDLDQTHSVYVDQWDWE 120
 5 IPNA FEVHSLAKWKRHTLAR F E EKL V+MKALRPDESDLD THSVYVDQWDWE 120

Sbjct: 61 CIPNAFEVHSLAKWKRHTLARFNFKEDEGLVHMALRPDESDLDQTHSVYVDQWDWE 120

Query: 121 KVIPOGKRNLAYLKETVETIYKIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDL 180
 KVIP+GARN AYLKETV +IY+ IRLTELAVEAR+DI ++LPK+ITF+H+E+LV +YPDL

Sbjct: 121 KVIPGGRNRNAYLKETVNSIYAIRLTELAVEARFDIPILPKQITFVHSEDLVKIRYPLD 180

Query: 181 TPKEKRNATIKRFGAVFLIGIGVLPGKPHDGRAPDYDWTITETENGYHGLNGDILVWN 240
 + KERENAI KE+GAVFLIGIG L DKGPHDGRAPDYDWTTE+ENGY GLNGDILVWN

Sbjct: 181 SSKERENAI CKRYGAVFLIGIGGLSGKPHDGRAPDYDWTITETENGYHGLNGDILVWN 240

Query: 241 DQLGSAPFELSSMGIRVDEALKRQVEMTGDQDRGLGPDWIKSLNLGLPLTTIGGGIGQSRM 300
 DQGS APFELSSMGIRVDE AL+ QV +TGD+D L DWH+ LING PLTTIGGGIGQSR+

Sbjct: 241 DQLGSAPFELSSMGIRVDESAKLQVGLTGEDHLKMDWHQDLGLPLTTIGGGIGQSR 300

Query: 301 VMFLLRKHIGEVQTSVMPQEVDRSYDNIL 330
 M LLRK+HIGEVQ+SVMP+E+ + + NIL

Sbjct: 301 AMLLLRKHIGEVQSSVMPKEMLEFESNIL 330

An alignment of the GAS and GBS proteins is shown below.

Identities = 254/303 (83%), Positives = 280/303 (91%)

Query: 1 MIDKLEIVEVQGPILSQVGDGMQDNLSIEHPVSVKVINIPEAEFEVHSLAKWKRHTLA 60
 +I KL++VEVQGPILS+VGDGMQDNLS E+PVS V L IP A FEVHSLAKWKRHTLA

Sbjct: 23 LIANKLVEVQGPILSRVGDGMQDNLSSTENPVSVMVKIIPNATFEVHSLAKWKRHTLA 82

Query: 61 RFGFNEGBGLVHMALRPDESDLDQTHSVYVDQWDWEKVIPOGGRNLDYLKETVEIKYK 120
 RFGFNEGBGL V+MKALRPDESDLD THSVYVDQWDWEKVIPOG+RNL YLKETVE IYK

Sbjct: 83 RFGFNEGBGLVNMKALRPDESDLDQTHSVYVDQWDWEKVIPOGKRNLAYLKETVETIYK 142

Query: 121 AIRLTELAVEARFDIESILPKRITFIHTEELVEKYPDLSPKERENAIKRYGAVFLIGIG 180
 IRLTELAVEAR+DIE++LPK+ITFIHTEELV KYPDL+PKERENAI KE+GAVFLIGIG

Sbjct: 143 VIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDLTPKERENAITKRFGAVFLIGIG 202

Query: 181 GELADGKPHDGRAPDYDWTTPSENGFGLNGDILVWNOLGTAFELSSMGIRVDEALK 240
 G L DKGPHDGRAPDYDWTTE +ENG+ GLNGDILVWN+QLG+APFELSSMGIRVDE+ALK

Sbjct: 203 GVLFGKPHDGRAPDYDWTTEENGYHGLNGDILVWNQDQLGSAPFELSSMGIRVDEALK 262

Query: 241 RQVVLTDGDRLEFEWHKTLRGGFFLTTIGGGIGQSRMLFLRLXKHIGEVQSSVMPKEV 300
 RQV +TGD+DRL F+WHK+LL G FPLTTIGGGIGQSR+ MELLRK HIGEVQ+SVMP+EV

Sbjct: 263 RQVEMTGDQDRLGPDWIKSLNLGLPLTTIGGGIGQSRVMFLLRKHIGEVQTSVMPQEV 322

Query: 301 RDT 303
 RD+

Sbjct: 323 RDS 325

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2287

A DNA sequence (GBSx2421) was identified in *S. agalactiae* <SEQ ID 7055> which encodes the amino acid sequence <SEQ ID 7056>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm -- Certainty=0.3163 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2288

A DNA sequence (GBSx2422) was identified in *S.agalactiae* <SEQ ID 7057> which encodes the amino acid sequence <SEQ ID 7058>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9007> which encodes amino acid sequence <SEQ ID 9008> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
Identities = 30/97 (30%), Positives = 50/97 (50%), Gaps = 3/97 (3%)

Query: 50 KALVSKQSEATIPIGRPTCCYCRFLPKLLKSQATLHSKIYYLDSQNYKG-KRLKSF 108
      +A + ++ AT FIGR TC YCR F L A + IY+++S++ L+FF
Sbjct: 18 RAQEAIDKKETATPFGIKTKTCYCRKFPAGTSLGVVAETKAHLYFINSEASQNLDAQFR 77

Query: 109 KKHHTTTPVNLAHYQQGKMTKYLNVQGSQATPQQIQTF 145
      ++ I TVP H G++ + S + Q+I+ F
Sbjct: 78 SRYGIPTVPGFVHITGGQIN--VRCDSSMSAQBIKDF 112

```

SEQ ID 9008 (GBS134) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 2; MW 17kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 4; MW 42kDa).

GBS134-GST was purified as shown in Figure 204, lane 10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2289

A DNA sequence (GBSx2423) was identified in *S.agalactiae* <SEQ ID 7059> which encodes the amino acid sequence <SEQ ID 7060>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0735 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9603> which encodes amino acid sequence <SEQ ID 9604> was also identified.

-2570-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAH06309 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 78/178 (43%), Positives = 115/178 (63%), Gaps = 3/178 (1%)

5 Query: 3 MRVAGTFGGRPLKTLDGKTRPTTDKVKRAIFNMIGPFFEGSRVLDLPSGSGSLAEAI 62
MRV+AG G LK + G TRPTTDKVK AI FNMIGPFF+GQ LDL+ GSG L IEA+
Sbjct: 1 MRVIAGBQKGLTLKAVPGHKTRPTTDKVKRAIFNMIGPFFDGIGLDLYGSGSGSLAEAL 60

10 Query: 63 SRGMDQAVLVKDRRAQVVVIQENIAMTKSPQPOLLMQANRALEQLTQQ---FDLVLLD 119
SRG+++ + V++ +RA I++N++ + ++ + +A RAL+ LT + F V LD
Sbjct: 61 SRGVRMIFVDQOKFAIETIKNLKSHCGLEGRAEVYRNDKRALQVLTKRGIVFAYVFLD 120

Query: 120 PPYAKSEIVKQIQITMDSGLLGDDIMACITKESVDLPREIASPGIWNQKIYGISKVT 177
PPYAK+ I + I+ + GLL + ++ CE D+ LP++I K++ YG + +T
15 Sbjct: 121 PPYAKQTIKNDLAILANHGILLESGVVVVCNDRDITWLPDQIEVAVKKHKEYTGDITMT 178
```

There is also homology to SEQ ID 132.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 2290

A DNA sequence (GBSx2424) was identified in *S. galactiae* <SEQ ID 7061> which encodes the amino acid sequence <SEQ ID 7062>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4984 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96619 GB:AJ006310 hypothetical protein [Streptococcus pneumoniae bacteriophage M41]
Identities = 175/254 (68%), Positives = 219/254 (85%)

35 Query: 2 LRRHIYSMLESHXHLQPEIKYHOKTNLRKRVYTVFIEKVDVILADLKADAFPGIETG 61
L RR+Y ++ EI++HQ++NLKKNRVYTVF +EKV +L+DL LAD+FFG+ETG
Sbjct: 50 LARHLYSEPLHFYETKSEIRHQRSLNRKRVYTVFSTOKSQDQLSLDLHLADSPFGLETG 109

Query: 62 IEHSILDNDENGGRAYLRGAFLSTGTVREPDGSKYQLEIFSVYLDHAQDLANLAKKPMFLDA 121
I+ +IL ++E GRAYL GAFL+ G++R+P+SGKYQLEI SVYLDHAQ ++L+++P+LDA
40 Sbjct: 110 IDEATLSDEEAGRAYLGAFLANGSTRDPESKYQLEITSSVYLDHAQSGIASLLQQFLDA 169

Query: 122 KVIEHKGHGAVTYQKAEDIMDFLIVIDAMEARDAPEEIKRMIRETRNDINRANNVETANIA 181
KV+E K GAVTYLQ+AEEDIMDFLIVI AN+ARD FE +K++RETRND+NRANN ETANIA
45 Sbjct: 170 KVLRRKKGAVTYLQKAEDIMDFLIVIGAMQARDPERVKILRETRNDINRANNAETANIA 229

Query: 182 RTTASMKTTINNIKIMDTTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLTPELSEKGV 241
RT++ASMKTTINNI KI D +G + LP DL++VAQ+R+ HPDYSIQQ+ADEL TPL+ESGV
50 Sbjct: 230 RTVSAEMKTTINNIKIKIMDINGLENLFDVLQVEAQLRIQHPDYSIQQIADSLTPLTKSGV 289

Query: 242 NHRLRKINKIADEL 255
NHRLRKINKIADEL
Sbjct: 290 NHRLRKINKIADEL 303
```

55 There is also homology to SEQ ID 5540:

```
Identities = 186/254 (73%), Positives = 227/254 (89%)

Query: 2 LRRHIYSMLESHXHLQPEIKYHOKTNLRKRVYTVFIEKVDVILADLKADAFPGIETG 61
+ R+IYS++E+ + PEI+YHOKTNLRKRVYTV++E+ V+ ILADLKAD+FFG+ETG
```

-2571-

5
 10
 15

```

  Sbjct: 50 IARYIYSLIEDAYVIVPEIRYHQKTNLRKNRVYTVYVQGVETILADLEKLADSPFGLETG 109
  Query: 62 IEHSILINDNRGRAYLRGAFSLSTGVREPDGKYQLEIFSVYLDHAQDLANLMKKFMFLDA 121
    IE +L +D GR+YL+GAFL+ G++R+P+SGKYQLEI+SVYLDHAQDLA LM+KFMFLDA
  Sbjct: 110 IEPQVLSDDNAGRSYLGKAFIAGSIRDPESGKYQLEIYSVYLDHAQDLAQLAKQFMFLDA 169
  Query: 122 KVIEHKHGAVTYLQKAEDIMDFLIVDAMBARDAPEEIKIMIRETNDINRANNVETANIA 181
    K IEHK GAVTYLQKAEDIMDFLI+I AM ++ FE IK++RE RNDINRANN ETANIA
  Sbjct: 170 KTIIEHSGGAVTYLQKAEDIMDFLIIGAMSCKRDFEAIKILREARNINRANNMETANIA 229
  Query: 182 RTITASMKNTINNIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQIADSLRPLSKSGV 241
    +TI+ASMKNTINNIKIMDTIG ++LP +L+QVAQ+RV HPDYSIQQ+AD+LE P++KSGV
  Sbjct: 230 KTIISAMKNTINNIKIMDTIGRLSFLPIELQVQAQLRVKHPDYSIQQVADALEFPITKSGV 289
  Query: 242 NHRLRKINKIADDEL 255
    NHRLRKINKIAD+L
  Sbjct: 290 NHRLRKINKIADDL 303
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2291

A DNA sequence (GBSx2425) was identified in *S.agalactiae* <SEQ ID 7063> which encodes the amino acid sequence <SEQ ID 7064>. Analysis of this protein sequence reveals the following:

25
 30

```

  Possible site: 14
  >>> Seems to have no N-terminal signal sequence

  ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.0297 (Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
                bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2292

A DNA sequence (GBSx2428) was identified in *S.agalactiae* <SEQ ID 7065> which encodes the amino acid sequence <SEQ ID 7066>. Analysis of this protein sequence reveals the following:

40
 45

```

  Possible site: 31
  >>> Seems to have no N-terminal signal sequence

  ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.2706 (Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
                bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

The protein has homology with the following sequences in the GENPEPT database.

50

```

  >GP:CA54571 GB:A7066393 response regulator [Streptococcus pneumoniae]
  Identities = 139/190 (73%), Positives = 166/190 (87%)

  Query: 8 IKIVLVDDHEMVRIGLKSFLNLQADVVEVIGRASNGLSGIIKALELRPDVVVMDLVMPEND 67
    +KI+LVDDHEMVRIGLKS+ +IQ DVSV+GRASNG +GI ALRLRPDV+VMD+VMPEN+
  Sbjct: 1 MKILLVDDHEMVRIGLKSYPDLQDDVEVVGSRASNGSGQIDLALRLRPDVVMDLVMPENI 60
  
```

-2572-

Query: 68 GVEATLALLKDWPEAAILVLTSYLONEKIYPVIEAGAGYMLKTSSAAEILNAINRKV 127
 G++ATLA+LK+WPEA IL++TSYLONEKI PV++AGAKTYMLKTSSA E+L+A+ KVA G
 Sbjct: 61 GIDATLAILKEWPEAKILIVTSYLONEKIMPVLGAGAGYMLKTSSADELLHVAHSVAA 120

5 Query: 128 EQAIRENEVDKKIKAHDKCPALHEGLTARERDIINLAKGYDNQRIADELFIISLTKVTHV 187
 E AIE EV KK++ H LIR LTAREKD+L LAKTY+NQRIAD+LFIISLTKVTHV
 Sbjct: 121 ELAIEQEVSKKVEYTHNIMELHEGLTARERDVLQLLAKGYENQRIADDFIISLTKVTHV 180

10 Query: 188 SNILAKLEVS 197
 SNIL KL S
 Sbjct: 181 SNILAKLEVS 190

There is also high homology to SEQ ID 2996:

Identities = 158/198 (79%), Positives = 176/198 (88%), Gaps = 1/198 (0%)

15 Query: 5 MDKIKIVLVDHDMVRLGLKSPFLMLQADVSVIGRASNGLEGTKKALELRPDVVMDLVMP 64
 M KIK++LVDDHDMVR+GLKSPFLMLQAD++V+GRASNG EG+ AL L+PDV+VMDLVMP
 Sbjct: 3 MSKIKIVLVDHDMVRLGLKSPFLMLQADIDVGRASNGREGVDLALALKPDVLVMDLVMP 62

20 Query: 65 EMDGVEATLALLKDWPEAAILVLTSYLONEKIYPVIEAGAGYMLKTSSAAEILNAINRKV 124
 E+ GVEATL+LK W EA +LVLTSYLONEKIYPVI+AGAGYMLKTSSAAEILNAINRKV
 Sbjct: 63 ELGVEATLEVLKWKAEKVLVLTSYLONEKIYPVIDAGAGYMLKTSSAAEILNAINRKV 122

25 Query: 125 SRGQEATRENEVDKKIKAHDKCPALHEGLTARERDIINLAKGYDNQRIADELFIISLTKV 184
 S+GE AIE EYDKKIKAH+ P LIR LTARE DIL+LLAGYDNQ IADELFIISLTKV
 Sbjct: 123 SGKEAIEETEVVDKKIKAHDKCPALHEGLTAREYDILHLAKGYDNQTIADDFIISLTKV 182

Query: 185 THVENILGKLN-GSRSNS 201
 THVENIL KL G R++
 30 Sbjct: 183 THVENILAKLEVGDRQA 200

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2293

35 A DNA sequence (GBSx2429) was identified in *S. agalactiae* <SEQ ID 7067> which encodes the amino acid sequence <SEQ ID 7068>. This protein is predicted to be histidine kinase (narQ). Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3944 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45

The protein has homology with the following sequences in the GENPEPT database.

>QP:CAE54570 GB:A7006393 histidine kinase [Streptococcus pneumoniae]
 Identities = 32/55 (58%), Positives = 49/55 (88%)

50 Query: 1 MIDNGIGFMDSDVYDLSYGLKNIERDVEDLAGNLQILSQPGKGVAMDRLPLVNVQ 55
 ++DNGIGF + S+ DLSYGL+NI++RVED+AG +QLL+ P +G+A+DIR+EL+++
 Sbjct: 276 VVDNGIGFQLOSLDLSYGLKNIERKVEDMAGTVQLLTAPKQLAVDIRIPLLDK 330

There is also homology to SEQ ID 2992:

Identities = 44/59 (74%), Positives = 51/59 (85%)

55 Query: 1 MIDNGIGFMDSDVYDLSYGLKNIERDVEDLAGNLQILSQPGKGVAMDRLPLVNVQ 59
 MID++G+FMID V DLSYGLKNIERK VEDLAGNL L+SQ GRGV+MDIRL+V++
 Sbjct: 276 MIDGIGFMDQVRLDLSYGLKNIERVDNLGAGNLILSQNGKGV8MDIRLPIVKGDEB 334

-2574-

Sbjct: 63 GKPRFBQQKKDKKARKWQKIVVMKEVRLSPITIDEIDFDTKLRNARKPLEK3DKVKCSIRF 122

Query: 127 KGRMITTHKEIGAKVLAEPAEATQDIAIEQRAKMDGRQMFLAPIDPK 175
KGR ITHKEIG KVL EA+A +D+ IEQR KMDGR MP+ LAE+ +K

5 Sbjct: 123 NGRALITHKEIGQKVLDRFAKACEDLCTIEQRKMDGRSMFVLAFLPEEK 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7075> which encodes the amino acid sequence <SEQ ID 7076>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2247 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

15 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 167/176 (94%), Positives = 173/176 (97%)

20 Query: 1 MKIIAKKDLFINDEIRVREVLVGLSEGLGIKPLSEAQAIAADDNDVLVLIQPQATPFV 60

+KIIAKKDLFINDEIRVREVLVGLSEGLGIKPLSEAQA++AD +NDVLVLIQPQA PFV

Sbjct: 1 VKIIAKKDLFINDEIRVREVLVGLSEGLGIKPLSEAQLADASNVDLVLIQPQAFV 60

25 Query: 61 AKIMDYGKFKFPEYQKQKQKRRKQSVVTVKEVRLSPVLDKQDFETKLRGRKFLKGNKV 120

AK+MDYGRFPEYQKQKQKRRKQSVVTVKEVRLSPVLDKQDFETKLRGRKFLKGNKV

Sbjct: 61 AKIMDYGKFKFPEYQKQKQKRRKQSVVTVKEVRLSPVLDKQDFETKLRGRKFLKGNKV 120

Query: 121 KVSIRFKGRMITTHKEIGAKVLAEPAEATQDIAIEQRAKMDGRQMFLAPIDPK 176

KVSIRFKGRMITTHKEIGAKVLA+PAEATQDIAIEQRAKMDGRQMFLAPIDPK

30 Sbjct: 121 KVSIRFKGRMITTHKEIGAKVLAEPAEATQDIAIEQRAKMDGRQMFLAPIDPK 176

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2296

35 A DNA sequence (GBSx2432) was identified in *S.agalactiae* <SEQ ID 7077> which encodes the amino acid sequence <SEQ ID 7078>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1807 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45308 GB:U81957 RNA polymerase beta' subunit [Streptococcus gordonii]

Identities = 262/286 (91%), Positives = 276/286 (95%)

50 Query: 1 MPAKVVIKAGVEVXIRSVFTCTNRHGVCRHCYGINLATGDAVEVGRAVGTIAAQSIGEPG 60

MA +VV AGV EV IRSV TCTNRHGVCRHCYGINLATGDAVEVGRAVGTIAAQSIGEPG

Sbjct: 122 MARQVNVAGVTEVTIRSVLTCNTRHGVCRHCYGINLATGDAVEVGRAVGTIAAQSIGEPG 181

Query: 61 TQLTMRFTFTGVSASNDITQGLPRIQRIPEARNPKEGAVITEVKGVEVAIEDSSRTTK 120

TQLTMRFTFTGVSASNDITQGLPR+QRIPEARNPKEGAVITEVKGVEVAIED+STRTTK

55 Sbjct: 182 TQLTMRFTFTGVSASNDITQGLPRVQRIPEARNPKEGAVITEVKGVEVAIEDASTRTK 241

Query: 121 KVFVKGQVTEGSEYVVPPTARMKVEVD+V+RGAAITGSIQPK IL VRD LSVETPYLLA 180

KVFVKGQVTEGSEYVVPPTARMKVEVD+V+RGAAITGSIQPK IL VRD LSVETPYLLA

Sbjct: 242 KVFVKGQVTEGSEYVVPPTARMKVEVDV+RGAAITGSIQPKHLLAVRDLVSVETPYLLA 301

-2576-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2299

A DNA sequence (GBSx2436) was identified in *S.agalactiae* <SEQ ID 7083> which encodes the amino acid sequence <SEQ ID 7084>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3015 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA18706 GB:U38906 Structural protein [Bacteriophage rit]
Identities = 57/127 (44%), Positives = 83/127 (64%)

Query: 5 IKAGTLKPKLVTREIMSKVKGBSTLAKLGGQTPIPPNGVQVFNLDGNAQIVGEGEQL 64
+ GILF P LVT+++SKV G S++A+LS Q PIPFNG + F F +D +V E +K
Sbjct: 3 LNKGTLDPTLTDLISKVAGKSSIALRLSAQKPIPPNGEKVFTFTDGEIDVVAESGKKT 62

Query: 65 GNTAKVTSKIIKPLKFVYQARMTDEFKYASSEKRLNFKHYADGPAKMGASAFDIAIHG 124
+ + + P+K Y AR++DEF YAS+E+++N L+ + DGPAKK+A D+ A HG
Sbjct: 63 HGGVTLPQTMVPIKVEYGARISDEFMYASDEEKINILQEFNDGPAKVARGIDLMAFHG 122

Query: 125 LEPRIMT 131
+ PR T
Sbjct: 123 VNRLLGT 129

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2300

A DNA sequence (GBSx2439) was identified in *S.agalactiae* <SEQ ID 7085> which encodes the amino acid sequence <SEQ ID 7086>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1892 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2301

A DNA sequence (GBSx2440) was identified in *S.agalactiae* <SEQ ID 7087> which encodes the amino acid sequence <SEQ ID 7088>. Analysis of this protein sequence reveals the following:

-2577-

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2227(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 10 vaccines or diagnostics.

Example 2302

A DNA sequence (GBSx2441) was identified in *S.agalactiae* <SEQ ID 7089> which encodes the amino acid sequence <SEQ ID 7090>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

15 Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.2948(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9319> which encodes amino acid sequence <SEQ ID 9320> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
 bacteriophage MW1]
 Identities = 84/238 (35%), Positives = 137/238 (57%), Gaps = 8/238 (3%)
 30 Query: 1 MTLDKNSSQPKKAGLILQKIEDRIAIKNHSEMTYGLKKRYLQWIFTVKDSITGRYL 60
 +T++K + Q+ +A ++LOEKI +L+ + +T+ E+ + K W TVK+STK
 Sbjct: 30 VTMRKKTTPQARNQAILLQKINKLSTKQVESITFEIYNLFYKSWAQTVKSTKGNCK 89
 35 Query: 61 VSDSHIATVLEDDTIINKLTKRDIRLIIDKLLKNSYHVHTKCRKRLHAIFS YAIQMDYM 120
 D + V+P DTI+ L +R ++ I+K+++ N Y K R RL IP+YA+Q Y+
 Sbjct: 90 SVDKIKKKEVIPSDTILANLDRFLQEAIEKIIBSGYITACKVRHRLGIFNYAVQSYXI 149
 Query: 121 TSNPTENVLVP-KPK--DDYKPEKVLVLYLTSNEV--YDLNRMIDNDEQLADIVLWFL 174
 -N + +P KPK ++ ++ +LT E+ D+ NR Q AD+VL + L
 40 Sbjct: 150 ENNEVDYTTIPQPKPTLSELEKKGNNFLTWQRIKALVDVLR--EYHQKYADMVLVLT 207
 Query: 175 TGVRYGELSCLTYDKIDFENKEILINATYDFTNRITTTTKKSTIRKISVSDNILDIV 232
 TG+RYGEL+ L IDFEN +I I +D + T XT S R I VS++++ +
 Sbjct: 208 TGMRYGELTALQKNIDFENNKIEITGNFDSVNIKIKLPKTNISRTIKVSEVISAL 265
 45

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 2303

50 A DNA sequence (GBSx2444) was identified in *S.agalactiae* <SEQ ID 7091> which encodes the amino acid sequence <SEQ ID 7092>. Analysis of this protein sequence reveals the following:

possible site: 50
 >>> Seems to have no N-terminal signal sequence

-2578-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2518 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 4212:

Identities = 92/144 (63%), Positives = 118/144 (81%), Gaps = 1/144 (0%)

10 Query: 1 MPKYSLFELENGRRRLASAGELQKGNELALFTQPMKFLYLASRYNESGKGPREEIKKQE 60
 +PKYSLFELENGR+R+LASAGELQKGNELALP+++ FLYLAS Y + KQ PE+ E+KQ
 Sbjct: 1198 LPKYSLFELENGRKRMLASAGELQKGNELALPISKYVNFLYLASHYEKLKGPSPEDEKQQL 1257

15 Query: 61 FVNQHVSYTDDILQLINDFSKRVILADANLEKINKLYQDNKSNISVDELANNINLFTFT 120
 FV QH Y D+I++ I++FSKRVILADANL+K+ Y +++ + E A NII+LFT T
 Sbjct: 1258 FVQHKHLYDEIIRQISEFSKRVILADANLQKLSAYNKHDKK-PIREQAENIIHLFTLT 1316

20 Query: 121 SLGAPAAPKFFDKIVDRKRYTSTQ 144
 +LGAAPAPK+FD +DKRYTST+
 Sbjct: 1317 NLGAPAPAPKYFDTTIDRKRYTSTK 1340

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2304

A DNA sequence (GBSx2445) was identified in *S. agalactiae* <SEQ ID 7093> which encodes the amino acid sequence <SEQ ID 7094>. This protein is predicted to be (). Analysis of this protein sequence reveals the following:

Possible site: 48

30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.57 Transmembrane 239 - 255 (236 - 256)

----- Final Results -----

35 bacterial membrane --- Certainty=0.2826 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]
 Identities = 88/257 (34%), Positives = 129/257 (49%), Gaps = 11/257 (4%)

45 Query: 1 MARLGADFYSLVTDLQKDGFTKFFQQTGVFLKKDESQESLFPALADKRRLSPGLIGD 60
 +A+ GA +X L+ L+KDG Y++ G + D S+L+ + A KRR ++P IGD
 Sbjct: 61 LANGSARYYKDLIHQLEKDGESDTYKRVGALS IHTDASKDKMERAYKRRDEAPSTIG 120

Query: 61 LQILNKSEANTHPPEL-DGYEQLLYASGGARVEGADLTRILLAS--GYNVIKDEVHP- 115
 + L+ SE FP L DGYE ++ SG ARV G L R LL A+ G VIK
 Sbjct: 121 IYKLSAESTKLLFPILADGYES-VHLSGAARVNGRALCKSLLSAARKRGATVIGKNASLL 179

50 Query: 116 ----TITDNGFRVQSIDFDKVLASGAWLAKILDEHNYQVIVRFQKQLRDYYFSNINTG 171
 T+T + D +++ +GWN +IL V QK Q+ + ++ +TG
 Sbjct: 180 FENGFTVGVQTDTKQPAADAVIVTGAWANELLKPLGIHPCVVSFQKAQINHPFEMTDATG 239

55 Query: 172 KYFVMPSEGLDIIIPFLNGKVSVGASHENIMAF-DLVIDFKVLDKFSEAIGYFPOLKKQ 230
 +PVVVP + I+ FDNQ+ GA+HEND DL + + +A+ P L
 Sbjct: 240 SWFVVPFSDQYILSPDNRIYAGATHENDAGLDLRLVTAGQGHVLSKALVAAGLADA 299

Query: 231 IRLLRVVEFVPIQVIFL 247
 + RV F P FL

-2579-

Subjct: 300 AAVETRVGFRPFTPGFL 316

There is also homology to SEQ ID 2656.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2305

A DNA sequence (GBSx2446) was identified in *S.galactiae* <SEQ ID 7095> which encodes the amino acid sequence <SEQ ID 7096>. Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2572 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9315> which encodes amino acid sequence <SEQ ID 9316> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC00337 GP:AF008220 YtqI [Bacillus subtilis]
Identities = 119/256 (46%), Positives = 174/256 (67%), Gaps = 3/256 (1%)

Query: 6 QILDRIKEYDTIIIRHMRPFDALGSGQIGLRDIIIRNFPFKKKVLATGFDEPTLAWIAK 65
+++ I YDTII+HRH+REFDFA GSG GL +I+R +P+K + A G EP+L+++ +
Sbjct: 4 ELIRTSISLYDTIILHRHVRFPDAYSQOGLTEILRETYPERKIFPAVGTPEPSLSFLYSL 63

Query: 66 DQVTDQYQDALVVVDTANTPRIDDERYKKGDFLIKIDHHPNDVYGDLSYVDTNASSA 125
D+V ++ Y+GALV+V DTAN RIDD+RY G L+KIDHHPM++ YGDL +VDT+ASS
Sbjct: 64 DEVINETYEGALVIVCDTANQERIDQRYPSGAKIMKINHHPNEDFYGDLAVDTSSAV 123

Query: 126 SEIVDFAL---SCDLLSTSAARVLYNGIVGDTGRFLYPATTSKTLKIAKLRFPDFD 182
SE++ + L L+T AA ++Y GIVGDTGRFL+P TT KILK A +L ++ F
Sbjct: 124 SEMIYELYLEGKEHGWKLNKAAELIYAGIVGDTGRFLPNTITKILKYAGELIQYFPSS 183

Query: 183 SAMARQMSDFPFKIAKLGQFIFELKIDKNGAACVLTQEDLKRFVDTAETAIVGVFG 242
S + Q+ + KL GFIP+ + + NGA V + ++ L++F T +E + +VG G
Sbjct: 184 SELFNQLYETKLNVVKLGKGFIPQNVLSSENGAASVFIKIDTLERFGTASEASQVLGTLG 243

Query: 243 KIDIVESWAIFVKQSD 258
I + +W FV++ D
Sbjct: 244 NISGIRAWFFVERD 259

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7097> which encodes the amino acid sequence <SEQ ID 7098>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2584 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 180/256 (70%), Positives = 215/256 (83%)

Query: 4 FQILDRIKEYDTIIIRHMRPFDALGSGQIGLRDIIIRNFPFKKKVLATGFDEPTLAWIA 63

```

-2580-

- F+ ILDKIK + TIIHRH POPDALGSQ GL++II NFP KKVL TGPDEP+LAMI+
 Sbjct: 5 FETILDKIKAHQITIIHRHQNPDPAIGSQAGLKEIIAQNFPOKKVIMTGFDEPSLAWIS 64
- Query: 64 MDQVTDQDTQALVVVTTANTPRIDDERYKKGDFLIKIDHHPNDEVGDLSTVDTNAG 123
 +MDQVTD+DY+ ALV++TDTAN PRIDDERY G LIKIDHHPND+VYGD YVDT+AS
 Sbjct: 65 QNDQVTDKDYKALVITITDTANRPRIDDERYTLGKCLIKIDHHPNDVYGDFFVYVDTNAG 124
- Query: 124 SASEIVTDFALSCDLLISTSAARVLNNGIVGDTGRFLYPRATTSKTLKIASLRFFDFDS 183
 SASEI+ DFA S +L LS AA++LF GIWGDTRFLY +TTSKTL IAS+LR F+DF+
 Sbjct: 125 SASEIADFAPSQNLTLSDKAALKLTYGIVGDTGRFLYASTTSKTLIASQLRHFFDFDA 184
- Query: 184 AMARQMDSPFPFKIAKLGQFIFRQLKIDKNGAACVTLTQRDAKRFVDIADTAATVGVGPK 243
 A++SQMDSPF KIAKLQ ++FE L ID++GAA V ++QE LK FVDT AE++AIV PGK
 Sbjct: 185 AISRQMDSPFLKIAKLGSTVFEHLITIDBSGAAYVLVSQETLKRFVDVTLAESSALVCAFGK 244
- Query: 244 IDIVESWALFVYKQSDG 259
 ID V++WALFV+ +DG
 Sbjct: 245 IDNVQWALFVFLITDG 260
- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2306

A DNA sequence (GBSx2447) was identified in *S. agalactiae* <SEQ ID 7099> which encodes the amino acid sequence <SEQ ID 7100>. Analysis of this protein sequence reveals the following:

- Possible site: 26
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.1846 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- >GP:CAB42949 GB:AL049863 putative adenosine deaminase [Streptomyces
 coelicolor A3(2)]
 Identities = 123/343 (35%), Positives = 175/343 (50%), Gaps = 26/343 (7%)
- Query: 6 LKELAKASLHCHLGGSLPAIRKIANMADIILPSSDK-KLRKYVIAPAQTSLVDYIKT 64
 L+ L KA LH HLDG L + +LA LP++D EL + A + LV Y+ T
 Sbjct: 11 LRRLPKAVLHDLGGLRPATVVELARSVGHTLPITDPDLGAWYVYEAANSGLVRYAT 70
- Query: 65 FEFIRFLQTKSALRFAAYDVARQALENVIYIEIRFAPELSMDRGLTASDVTLAVLEGL 124
 FE ++Q +E L AA + A + V+Y E+R+AFEL+ GL+ + V V EGL
 Sbjct: 71 FEHTLAVMGNRSGLLRAAEYVYLLDLADGVVYGEVRYAPELNTREGGLSMREVVTVQGL 130
- Query: 125 ADNQKEFNIVAR-----ALVCGMRQSGSHKTKDIKIHVDLA----PKGLNGFDFAGDEF 175
 A + + L+CGMR D ++ DLA G+VGFD AG E
 Sbjct: 131 ATGAKAAGAGTTPVRVGYLLCGMRMF-----DRVRKAADLAVAPRAGVVGFDLAGED 184
- Query: 176 SYFTSDSLVDLIQVNRSGYPWTLHAGCGCAKHIADSLAL-GLKRMGHVITALT----- 227
 +P +D + ++R P T+HAGE I +L + G +R+GH +T
 Sbjct: 185 GFFPADHLDAPEHLRRENVPFTITHGAGHGLPSLHQAQVCGAGSIGHGVRITIDDPDLA 244
- Query: 228 -GQDLIKRFVSEDAVA-EMCLTSNLQTKAASSIQSPFYQELYDAGKITINIDNRVSD 285
 G+ + +V+ +A EMC TSNLQT AA+SI P L D G ++T+NDNRVS
 Sbjct: 245 AGKLGRLAANVRDRIALEMCPNTSLNLTGAATSIASHPTALDGLGFRVTVINIDNRLVSG 304
- Query: 286 TNLTKYSGLPFTYPTGKTKIEDFLVFNKNAVKASPTSDSSKOTLL 328
 T +T+E SL V G +ED NN+K+I+ F E+ L+
 Sbjct: 305 TITIREMSLLVBQAGNSVEDLRTVTYVALKSAFVFPFDRITALI 347

-2581-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2307

- 5 A DNA sequence (GBSx2448) was identified in *Sagalactiae* <SEQ ID 7101> which encodes the amino acid sequence <SEQ ID 7102>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2642 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 A related GBS nucleic acid sequence <SEQ ID 9639> which encodes amino acid sequence <SEQ ID 9640> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CB13290 GB:Z99111 similar to sulfite reductase [Bacillus subtilis]
Identities = 63/146 (43%), Positives = 87/146 (59%), Gaps = 1/146 (0%)

- 20 Query: 5 MALAKIVYASMTGNTSEIADIADKLRDLGLDVEVECTMVDAAD-FEDADIAIVATYTY 63
NA +VTA+H+GNTIE +AD++ L++ +V+ E +D A F D D I+ TTT+
Sbjct: 1 MAKILLVIATNINSGNTEAMADLIEKGLQEALEVDREFEAMDIDDAQLFTDHYHVGINTYTW 60
- 25 Query: 64 GEGDLFDEIVDFYEDLAEVDLSGKGVGVGGSDTFYDFPKSVDEFEAQFALTGAQKQAD 123
GEGDLFDE +D ED+ E+D SGK V GSGDT Y++FC +VD EA+ G
Sbjct: 61 GEGDLFDEIVDFYEDLAEVDLSGKGVGVGGSDTFAYEFCGAVDTLEAKIKERGSDIVLP 120
- 30 Query: 124 CVKVDLAAEDEDINLEAFAPAEIASK 149
VK++ E E+ E L F + A K
Sbjct: 121 SVKININPGEIEEELINFGROQFAKK 146

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7103> which encodes the amino acid sequence <SEQ ID 7104>. Analysis of this protein sequence reveals the following:

- 35 Possible site: 14
>>> Seems to have no N-terminal signal sequence
- 40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1641 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- Identities = 116/147 (78%), Positives = 136/147 (91%)
- 45 Query: 5 MALAKIVYASMTGNTSEIADIADKLRDLGLDVEVECTMVDAAD-FEDADIAIVATYTYG 64
MALAKIVYASMTGNTSEIADIADKLRDLGLDVEVECTMVDAAD-FEDADIAIVATYTYG 64
Sbjct: 1 MALAKIVYASMTGNTSEIADIADKLRDLGLDVEVECTMVDAAD-FEDADIAIVATYTYG 60
- 50 Query: 65 DGLDFDEIVDFYEDLAEVDLSGKGVGVGGSDTFYDFPKSVDEFEAQFALTGAQKQAD 124
DGLDFDEIVDFYEDL ++DL GK+YGVVGGSDTFYDFPKSVDEFEAQFALTGAQKQAD 124
Sbjct: 61 DGLDFDEIVDFYEDLAEVDLSGKGVGVGGSDTFYDFPKSVDEFEAQFALTGAQKQAD 120
- 55 Query: 125 VKVDLAAEDEDINLEAFAPAEIASKIN 151
VKVDLAAEDEDI+ LEAFAPAEI++ +N
Sbjct: 121 VKVDLAAEDEDINLEAFAPAEIASKIN 147

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2308

- 5 A DNA sequence (GBSx2449) was identified in *S. agalactiae* <SEQ ID 7105> which encodes the amino acid sequence <SEQ ID 7106>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3568 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AE98234 GB:U67480 chorismate mutase/prephenate dehydratase
(phaA) [Methanococcus jamaeschi]
Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

20 Query: 2 ELESIHQSIDIDQQLVSLSTRWGLILEVIAFKKKHRLPVLNNRREVLNNVLKKVQN 61
+L KIR++IDEID +++ L+ R L +V K + +P+ D RE + + + K +
Sbjct: 4 KLASIRKKIDEIDNKIKLIAERNSLAKDVAEIKNLGIPINDFEREKYIYDRIKLCKE 63

25 Query: 62 HQFDDVIRATPKDIMTS-SRVYQKE 85
R D+ I I + E ++ QK+
Sbjct: 64 HNVDENIGIKIPQILIEHNKALQKQ 88

There is also homology to SEQ ID 1568.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2309

A DNA sequence (GBSx2450) was identified in *S. agalactiae* <SEQ ID 7107> which encodes the amino acid sequence <SEQ ID 7108>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

35 possible site: 23
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1828 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:ANC34413 GB:AF158600 putative minor structural protein
[Streptococcus thermophilus bacteriophage Sfill]
Identities = 39/65 (60%), Positives = 54/65 (83%)

50 Query: 1 MEVEITDSQVIMSTGLKOLKARHAYPAITYEVDGYVDLSQDVVRIQDDGYEPPLIITARV 60
ME++IDS++VL+ST D++L+ YPAITYEVLG++DL++GD V+IQD G+ P L+L ARV
Sbjct: 707 MEIDTDSBDVLI STALRNLRKFCYPAITYEVDGFLDLDIGITVKIQDGTGSPMIMLEARV 766

Query: 61 VEQDI 65
EQ I
Sbjct: 767 SEQQI 771

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2310

A DNA sequence (GBSx2451) was identified in *S.agalactiae* <SEQ ID 7109> which encodes the amino acid sequence <SEQ ID 7110>. This protein is predicted to be phosphomethylpyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2051(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:ANC22074 GB:U32725 phosphomethylpyrimidine kinase (thiD)
 [Haemophilus influenzae Rd]
 Identities = 29/78 (37%), Positives = 48/78 (61%), Gaps = 2/78 (2%)
 Query: 4 RNVLAIISGNDIFSGGGLHDLATYVYVNLHGFVAVTCLTMSDKG-FEVIPIEASILKQ 62
 + VL I+G+D G G+ ADL T+ + + G A+T +TA + G E+ + FI ++ Q
 Sbjct: 5 KQVLTITGSSGGGAGIQAQLKTFQMRSVFGTSAITAVTAQNTLGVDFIHLPLKTIQ 64
 Query: 63 LESLK-DVEFGSIKLL 79
 LE+K D + S K+G+L
 Sbjct: 65 LEAVNDFQIASCKIOML 82

30 There is also homology to SEQ ID 4408.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2311

A DNA sequence (GBSx2452) was identified in *S.agalactiae* <SEQ ID 7111> which encodes the amino acid sequence <SEQ ID 7112>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -7.43 Transmembrane 109 - 125 (102 - 129)
 INTEGRAL Likelihood = -1.28 Transmembrane 84 - 100 (84 - 100)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA22372 GB:AL034446 putative transmembrane protein
 [Streptomyces coelicolor A3(2)]
 Identities = 25/93 (26%), Positives = 43/93 (45%), Gaps = 1/93 (1%)
 Query: 62 SASVEILCRGWLIPVATKYKIVSVSISIFFGLLHSAHNVSLISIFNLCL-FGLFLS 120
 +A+ E++ RG L + +++ ++ + FGL+H N +L + + G L+
 Sbjct: 143 ANTEVVVRGVLFRILIEHHIGTYIALGLTGLVFGMLHLLNEDATLNGALALAIRAGFMIA 202

-2584-

Query: 121 LVVILKGINIWACGIIHGANVCVGSVFGIEVSG 153
 N+W G+H NN G VF VSG
 Sbjct: 203 AAYAATRNLM/LTIGVHFQWNPAGGVSTTVVSG 235

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2312

A DNA sequence (GBSx2453) was identified in *S.agalactiae* <SEQ ID 7113> which encodes the amino acid sequence <SEQ ID 7114>. This protein is predicted to be pppl protein. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5796 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA10712 (B:AJ132604) pppl protein [Lactococcus lactis]
 Identities = 38/64 (59%), Positives = 51/64 (79%)
 Query: 1 MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHHRAGNIASENTVTDLGSDDA 60
 ME S+L+DIG +RS NQD++ + N+AG L+LADGMGGH+RGN+AS++TV DLG W+
 Sbjct: 1 MEISILSDIGSRSTNQDFVSTVYVNRAGYQLFLADGMGGHHRAGNVASKLTIVEDLGKLS 60
 Query: 61 ETDF 64
 ET F
 Sbjct: 61 ETFF 64

There is also homology to SEQ ID 3022:

Identities = 58/74 (78%), Positives = 69/74 (92%)
 Query: 1 MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHHRAGNIASENTVTDLGSDDA 60
 M+ISL TDIGQ+RSNNQDFIN+FNK G+ L+ILADGMGGHHRAGNIASENTVTDLG +W
 Sbjct: 1 MKISLKTIDIGQRRSNNQDFINQFENKAGITLVILADGMGGHHRAGNIASENTVTDLGREVV 60
 Query: 61 ETDFSELSEIRDM 74
 +TDF+ELS+IRDM+
 Sbjct: 61 KTDFTLSQIRDNL 74

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2313

A DNA sequence (GBSx2454) was identified in *S.agalactiae* <SEQ ID 7115> which encodes the amino acid sequence <SEQ ID 7116>. This protein is predicted to be sunL protein. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

-2585-

bacterial cytoplasm --- Certainty=0.1631 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA10711 GB:A132604 sumL protein [Lactococcus lactis]
 Identities = 48/81 (59%), Positives = 67/81 (82%)

10 Query: 1 MSILSSVCQLRKGGIITYSTCTIPEENFQVIRKFLNHNPFQVLSHTQEDIVKRCG 60
 + IL+S +L+K GI+ YSTCTIP+ENP V+ +FLENHNPFQV+S+ + ++K GC
 Sbjct: 342 LEILNSASKSLKKGDMVYSTCTIPDENFVDVHEFLENHNPFQVLSNEKPEVIRKCG 401

15 Query: 61 ISISPEQYHTDGFPIQVVKRI 81
 + I+FE YHTDGFPI + K+I
 Sbjct: 402 LFIITPEMYHTDGFPIAKFKKI 422

There is also homology to SEQ ID 3018:

Identities = 64/82 (78%), Positives = 74/82 (90%)

20 Query: 1 MSILSSVCQLRKGGIITYSTCTIPEENFQVIRKFLNHNPFQVLSHTQEDIVKRCG 60
 + ILSSVCQLRKGGIITYSTCTIP+EN QVIR FL++HNPFQV+L+HTQ DIVK G
 Sbjct: 359 LEILSSVCQLRKGGIITYSTCTIPDENQVIRFLQSHNPFQVLSHTQEDIVKRGY 418

25 Query: 61 ISISPEQYHTDGFPIQVVKRI 82
 + I+PEQY TDGFFIQV+R+L
 Sbjct: 419 LIITPEQYHTDGFPIQVVRVL 440

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 2314

A DNA sequence (GBSx2455) was identified in *S. agalactiae* <SEQ ID 7117> which encodes the amino acid sequence <SEQ ID 7118>. This protein is predicted to be PTS permease for mannose subunit IIPMan.

Analysis of this protein sequence reveals the following:

Possible site: 53
 35 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.18 Transmembrane 32 - 48 (30 - 58)
 INTEGRAL Likelihood = -8.07 Transmembrane 127 - 143 (122 - 146)
 INTEGRAL Likelihood = -2.07 Transmembrane 56 - 72 (56 - 72)
 INTEGRAL Likelihood = -1.44 Transmembrane 87 - 103 (86 - 103)
 40 INTEGRAL Likelihood = -0.53 Transmembrane 105 - 121 (105 - 121)

----- Final Results -----
 bacterial membrane --- Certainty=0.4673 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF81084 GB:AF228498 AgaW [Escherichia coli]
 Identities = 38/122 (31%), Positives = 68/122 (55%), Gaps = 7/122 (5%)

50 Query: 25 KVPETKSIIRLTALAFVCSILNVLVSMRELISISFIIIGLVGSGPVNFSFVHHHPQNLN 64
 ++P T + L A +L L+++ +P I G+ + + PQ L+
 Sbjct: 126 KMPRTPLAALNAACNYLA-----LIALGNFYFLCAFLPIYFGARHAKTITDVLQRLI 178

55 Query: 85 NGLSAGGCLLPVAVGFAMLMKLLWINKLAFFVLLGFVLTAYLKLPAVAVAAAGVICVISS 144
 +GL AGG++EA+GFA+L+K++ N +++LGFV A+LKL +A+A + +I
 Sbjct: 179 DGLGVAGGIMPAIGFAVILLKNNKNNVYIYPIILGFVAAWMLKLPVLAACPALMAALIDL 238

Query: 145 QR 146

-2586-

R
 Sbjct: 239 LR 240

There is also homology to SEQ ID 1636:

5 Identities = 104/109 (95%), Positives = 108/109 (98%)

Query: 56 LIISISFIGILVGGSPVNSPVHHITPQNLNMGLSAAGGLPAVGFAMLMKLLNTNKLAVFY 115
 +I+SIISFIGILVGGSPVN+PV HIPQNLNMGLSAAGGLPAVGFAMLMKLLNTNKLAVFY
 10 Sbjct: 149 IIAISISFIGILVGGSPVNAFVHHITPQNLNMGLSAAGGLPAVGFAMLMKLLNTNKLAVFY 208

Query: 116 LLGFVLITAYLKLPAVAVALGAVICVISSQRIELDAITRGAISKQTTFF 164
 LLGFVLITAYLKLPAVAVALGAVICVISSQRIELDAITRGAISKQTTFF
 15 Sbjct: 209 LLGFVLITAYLKLPAVAVALGAVICVISSQRIELDAITRGAISKQTTFF 257

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2315

A DNA sequence (GBSx2456) was identified in *S.agalactiae* <SEQ ID 7119> which encodes the amino acid sequence <SEQ ID 7120>. Analysis of this protein sequence reveals the following:

20 Possible site: 50
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -8.12	Transmembrane	121 - 137 (118 - 144)
INTEGRAL	Likelihood = -5.52	Transmembrane	91 - 107 (89 - 111)
INTEGRAL	Likelihood = -5.20	Transmembrane	166 - 182 (162 - 192)

25 ----- Final Results -----

bacterial membrane	---	Certainty=0.4248 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15963 GB:Z99124 phosphotransferase system (PTS)
 beta-glucoside-specific enzyme IIABC component [Bacillus subtilis]
 Identities = 76/201 (37%), Positives = 122/201 (59%), Gaps = 3/201 (1%)

35 Query: 1 MIKALLALLVFKILTPSSQTYILLNLFDQGVFFPLPILIAITAAKQKLANPILALGTGV 60
 MIK L+AL + F + SQ +++L DG FYFLP+L+A++AA+K +NP +A
 Sbjct: 121 MIKQVALAVTFQNMAEKSOVHVILTAVGDFAFYPLPILLMSAARKFGSGNPYVAALIAA 180

40 Query: 61 MLLEPNVANLVASGKPVSLFHTIPPTLITNYASSVIPILIIICVQAYIEKYLKQIIPKSLR 120
 +LHP+ L+ +GKP+S F +P T Y+S+VIPI+L I + +Y+EK++ + SL+
 Sbjct: 181 ALLEPDLITALLGAGKPIIS-FIGLPVTAATYSSSTVIPILLISWIASYVEKNIDRFTASLK 239

45 Query: 121 LVLVPMILFLSMGILSFSILOPMQTTAGQYLAVITFLSKYASW-APAPLVGAFAPILIM 179
 L++VP L + L+ +GP+G I G+YL+ +L +A A FL G F+ ++IM
 Sbjct: 240 LLVVPTFTLLVVLPLTLITVGLGAILGEYLSSGVNYLFDHAGLVANIMFLAGTFS-LIIM 298

Query: 180 FGVHSGILAAIGITQLAKIGVD 200
 G+H + I +A+ G D
 50 Sbjct: 299 TGMHYAFVFTIMINLIAGNGHD 319

There is also homology to SEQ ID 2884.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2587-

Example 2316

A DNA sequence (GBSx2457) was identified in *S. agalactiae* <SEQ ID 7121> which encodes the amino acid sequence <SEQ ID 7122>. This protein is predicted to be glucose kinase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 54
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10    bacterial cytoplasm --- Certainty=0.1180(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB14416 GB:Z99116 glucose kinase [Bacillus subtilis]
    Identities = 32/57 (56%), Positives = 41/57 (71%)

    Query: 1   MVIGGGVSAAGEFLRSRVEKYFVTIFAPPQVKSTKIKIAELGNDAGIIGASLANQ 57
              +V+GGGVS AGE LRS+VEK F  AFP+  ++ I IA LGNDAG+IG A +A +
20  Sbjct: 258 IVLGGGVSRAGELLRSKVEKTRKCAFPRAQAADISIAALGNDAGVIGGAWIAKNE 314

```

There is also homology to SEQ ID 198. An alignment of the GAS and GBS proteins is shown below:

```

    Identities = 50/56 (89%), Positives = 53/56 (94%)

    Query: 1   MVIGGGVSAAGEFLRSRVEKYFVTIFAPPQVKSTKIKIAELGNDAGIIGASLANQ 56
              +VIGGGVSAAGEFLRSR+EKYFVTIF PPQV+ STIKIKIAELGNDAGIIGASLA Q
25  Sbjct: 264 VVIGGGVSAAGEFLRSRIEKYFVTIFPPQVRYSTKIKIAELGNDAGIIGASLARQ 319

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2317

A DNA sequence (GBSx2458) was identified in *S. agalactiae* <SEQ ID 7123> which encodes the amino acid sequence <SEQ ID 7124>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 19
    >>> Seems to have a cleavable N-term signal seq.

    ----- Final Results -----
    bacterial outside --- Certainty=0.3000(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

    The protein has homology with the following sequences in the GENPEPT database.

    >GP:CAB14385 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
    Identities = 37/86 (43%), Positives = 51/86 (59%)

45  Query: 3   MSVILITVILLAPVAMASWNYKVRRAAKFLONESFOKEMSRGQLIDIREAGAFHRKHIL 62
              MG +++++I AF+ + +Y +R K L E F+ + QLD+RE F HIL
    Sbjct: 1   MNNKIVLIFFPATITMYASVYVQQRIMCTLTETEEPRAGYRKAGLIDVREPNEFEGGHIL 60

    Query: 63  GARNIFASQFKVALSALRKDKPVLLY 88
50  GARNIF SQ K + +R DKVF LY
    Sbjct: 61  GARNIFLSQLKQRKNEIRTDKPVYLY 86

```

There is also homology to SEQ ID 202. An alignment of the GAS and GBS proteins is shown below:

```

    Identities = 51/108 (47%), Positives = 70/108 (64%)

```

55

-2588-

```

Query: 1  MEMSVIIITVILLAPVAMAGWNYWVRRAAKFLDMESPOKEMSRQQLIDIREAGAPHRKH 60
M    +++ ++L+ V + +WNT+ R+ AK +DNE+P+ M +GQLID+RE AF KH
Sbjct: 1  MSPITLILMLLLAGVIGYTYWNYPSFKIMAKQVNETPKDVMRQQLIDILREPAAFKTKH 60

5  Query: 61  ILGARNIPASQPKVALSALRKDKPVLVYDASRGQSIPRIVLLLRKEFF 108
ILGARN PA QF A+ LRKDKPVL+Y+ R Q V L+K F
Sbjct: 61  ILGARNFPAQQFDAAIKLRKDKPVLIIYENMRPQYRVPAVKKLGAGF 108

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2318

A DNA sequence (GBSx2459) was identified in *S.agalactiae* <SEQ ID 7125> which encodes the amino acid sequence <SEQ ID 7126>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1892 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2319

A DNA sequence (GBSx2460) was identified in *S.agalactiae* <SEQ ID 7127> which encodes the amino acid sequence <SEQ ID 7128>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3522 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2320

A DNA sequence (GBSx2461) was identified in *S.agalactiae* <SEQ ID 7129> which encodes the amino acid sequence <SEQ ID 7130>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2770 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-2589-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AA18708 GB:U38906 ORF33 [Bacteriophage rit]
Identities = 56/85 (65%), Positives = 66/85 (76%), Gaps = 1/85 (1%)

Query: 1 MNPFAITDDVILLNRQLSVDEIKRAELTSTVSDTLREASKVGKILDEMLITP-YFAT 59
M FAT DD+ +LNR L DE +RAE LLE VSD+LR EA KVG++L MI E P YFA+
Sbjct: 1 MNPFAITVDLITMLNRPLRGDKERRAEKILEIVSDSLREADKVGRLDYAMIAEKPSYFAS 60

Query: 60 VLKSVTVDIVARTLMTATQGEPMQ 84
V+KSVTVDIVARTLMT+T EPM+Q
Sbjct: 61 VLKSVTVDIVARTLMTSTQGEPMQ 85

There is also homology to SEQ ID 1432.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2321

A DNA sequence (GBSx2462) was identified in *S. agalactiae* <SEQ ID 7131> which encodes the amino acid sequence <SEQ ID 7132>. This protein is predicted to be regulatory protein TypA (typA). Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2238 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06351 GB:AP001516 GTP-binding protein TypA/BipA (tyrosine phosphorylated protein A) [Bacillus halodurans]
Identities = 175/237 (73%), Positives = 204/237 (85%), Gaps = 1/237 (0%)

Query: 1 MEDIFVGETVPTDAIEPLFVLRIDEPTLQMTFLVNNSPFAGREGKWTSRKVEERLAE 60
ME+I VGETV P D +PL+LRIDEPTLQMTFLVNNSPFAGREGK +TSRK+ERL AE
Sbjct: 281 MEINVGETVCFVNDHDFLILRIDEPTLQMTFLVNNSPFAGREGKHVTSRKLEERLAE 340

Query: 61 LQTDVSLRVDPDTPDKWTVSGRGELHLSILITMRBRYELQVSRPEVIIKEIDGVQCE 120
L+TDVSLRV+ TDSPD W VSGRGELHLSILIE MRBRYELQVS+FEVII+EIDGVQCE
Sbjct: 341 LETDVSRLVNTDTPDMVWVSGRGELHLSILIE MRBRYELQVSKPEVIIKEIDGVQCE 400

Query: 121 PFERVQIDTPREYQGAIIQSLSERGDMLDQMVGNGQIRLI FLIPARGLIGYSTFLSM 180
P ERVQID PREY GA+++SL ERKG+ML+M G+GQ RL F++PARGLIGY+TFPLS
Sbjct: 401 PFERVQIDPREYTGAVNESLGERGDMLEMTNTGSGVRLFMVARGLIGYTFEPLSQ 460

Query: 181 TRGYGIHNTFDQYLFVVOGEIGGRHGAALVSIENGKATTYSIMRIEERGNLSFVNP 237
TRGYGI+NH+FD Y PV G+GGR +G LVS+E GRAT Y I+++E+RG +V P P
Sbjct: 461 TRGYIINISFDQYQFVTPQGVGRGQVLSVMEIGKATQYGIQIVDGRGTI-FVSP 516

There is also homology to SEQ ID 206. An alignment of the GAS and GBS proteins is shown below:

Identities = 228/237 (96%), Positives = 233/237 (98%), Gaps = 1/237 (0%)

Query: 1 MEDIFVGETVPTDAIEPLFVLRIDEPTLQMTFLVNNSPFAGREGKWTSRKVEERLAE 60
MEDIFVGET+TPTD +E LP+LRIDEPTLQMTFLVNNSPFAGREGKWTSRKVEERLAE
Sbjct: 284 MEDIFVGETITPTDCVALLFLRIDEPTLQMTFLVNNSPFAGREGKWTSRKVEERLAE 343

Query: 61 LQTDVSLRVDPDTPDKWTVSGRGELHLSILITMRBRYELQVSRPEVIIKEIDGVQCE 120

-2590-

```

      LQTDVSLRVDPTDSPDKVTVSGRGELHLISILITMRREGYELQVSRPEVLIKIDGV+CE
      Sbjct: 344 LQTDVSLRVDPTDSPDKVTVSGRGELHLISILITMRREGYELQVSRPEVLIKIDGVKCE 403
5  Query: 121 PFERVQIDTPPEYQGAIIQSLSERKGDWLDQMVQNGQQLRIPLI PARGLIGYSTEFISM 180
      PFERVQIDTPPEYQGAIIQSLSERKGDWLDQMVQNGQQLRIPLI PARGLIGYSTEFISM
      Sbjct: 404 PFERVQIDTPPEYQGAIIQSLSERKGDWLDQMVQNGQQLRIPLI PARGLIGYSTEFISM 463
10 Query: 181 TRGIGIMNHITFDQILPVVQGEIGSRHGRGALVSIENGKATTYSIMRIEERGNLSFVNP 237
      TRGIGIMNHITFDQILPVVQGEIGSRHGRGALVSIENGKATTYSIMRIEERGN + FVNP
      Sbjct: 464 TRGIGIMNHITFDQILPVVQGEIGSRHGRGALVSIENGKATTYSIMRIEERGTI - FVNP 519

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2322

- 15 A DNA sequence (GBSx2464) was identified in *S. agalactiae* <SEQ ID 7133> which encodes the amino acid sequence <SEQ ID 7134>. This protein is predicted to be pseudouridine synthase family 1 protein (ruB). Analysis of this protein sequence reveals the following:

```

      Possible site: 34
      >>> Seems to have no N-terminal signal sequence
20  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1950 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
25  The protein has homology with the following sequences in the GENPEPT database.
      >GP:CA814248 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
      Identities = 59/105 (56%), Positives = 85/105 (80%)
30  Query: 5 VKERIVPVGRLEWDITGLLILTNDDGDFDTDMIHPRNEIDKVVYLRVKGIAIKENLRPLTR 64
      + RIVP+GRLD+DT+GLL+LTNDG+P +K++HP+ EIDK Y+A+VKEGI KE LR L R
      Sbjct: 91 IPQRIYPVGRLEWDITGLLILTNDDGDFDTDMIHPRNEIDKVVYLRVKGIPKELLRKLER 150
35  Query: 65 GVVIDGKTKKPARVYTIKVDHEKNRSVVELTIEHGRNHQVKNMFE 109
      G+ ++ NT PA+ ++ +D +K S++LTIHGRN QV++MFE
      Sbjct: 151 GIRLEEGKTAPAKAKLLSLDKKQTSITQLTIEHGRNHQVKNMFE 195

```

There is also homology to SEQ ID 4728:

```

      Identities = 96/109 (88%), Positives = 106/109 (97%)
40  Query: 1 MLFPQVKERIYPVGRLEWDITGLLILTNDDGDFDTDMIHPRNEIDKVVYLRVKGIAIKENLR 60
      +LPQVKERIYPVGRLEWDIT+G+LILTNDDGDFDT MIHPRNEIDRVYLRVKGIAIKENLR
      Sbjct: 94 LLFPQVKERIYPVGRLEWDITGVLILTNDDGDFDTMIHPRNEIDRVYLRVKGIAIKENLR 153
45  Query: 61 PLTRGVVIDGKTKKPARVYTIKVDHEKNRSVVELTIEHGRNHQVKNMFE 109
      PLTRG+VIDGKTKKPARV I+A+V+ +K+R+VELTIEHGRNHQVKNMFE
      Sbjct: 154 PLTRGVVIDGKTKKPARVNTVVRVADKRSIVELTIEHGRNHQVKNMFE 202

```

- 50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2323

A DNA sequence (GBSx2466) was identified in *S. agalactiae* <SEQ ID 7135> which encodes the amino acid sequence <SEQ ID 7136>. This protein is predicted to be L-rubulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

-2591-

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2827(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AA45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase
 [Bacillus stearothermophilus]
 Identities = 68/103 (66%), Positives = 82/103 (79%)

15 Query: 2 QEMRRVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVYDYLTPENMVVTDLE 61
 +E+++ V EAN LF + LV FTWGNVS +DRS GL+VIKPSGV YD+LT ++MVV DL
 Sbjct: 3 EELKQAVLEANLQPLQYRLVFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDLT 62

Query: 62 GNIVEGDLNPSSDLFTHVLQYKAWPEVGGIVHSTHSTAVGWAQ 104
 GN+VEGDL PSSD PTH+ LYK +P +GGIVHSTH A WAQ

20 Sbjct: 63 GNVVEGDLNPSSDTFTHLMLYKQFPFGIGGIVHSTHSTAVGWAQ 105

There is also homology to SEQ ID 4600:

Identities = 93/103 (90%), Positives = 96/103 (92%)

25 Query: 2 QEMRRVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVYDYLTPENMVVTDLE 61
 QEMRRVC ANKSLF H LKFTWGNVSEV RE G IVIKPSGVYDYLTPENMVVTDL+
 Sbjct: 6 QEMRRVCAANKSLPQHLVFTWGNVSEVCRBLORIVIKPSGVYDYLTPENMVVTDLE 65

Query: 62 GNIVEGDLNPSSDLFTHVLQYKAWPEVGGIVHSTHSTAVGWAQ 104
 GN+VEGDLNPSSDLFTHV+LYKAWPEVGGIVHSTHSTAVGWAQ

30 Sbjct: 66 GNVVEGDLNPSSDLFTHVELYKAWPEVGGIVHSTHSTAVGWAQ 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 2324

A DNA sequence (GBSx2467) was identified in *S. agalactiae* <SEQ ID 7137> which encodes the amino acid sequence <SEQ ID 7138>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3452(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:ARG05712 GB:AB004658 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 141/200 (70%), Positives = 162/200 (80%), Gaps = 1/200 (0%)

Query: 10 LSLGIDYETLANRFRPIPREISAGNVREKARALFYPIEWLKKAGFGAVRVPSSEYGGAG 69
 LS G DYR LA RFRPIF I+ G VERB+ R LP+R I WLK+AGHFGAVRVP E+GGAG
 Sbjct: 14 LSEADYELLAQFRPIFARIAGAVRERQRLPHEAIAWLKQAGFGAVRVPREHGGAG 73

Query: 70 ASIGQLFQLLIELARADSNIPQALRAHFAFVEDRLNAPPGVDRITWHARFVAGDLVGNW 129
 AS+ QL QLLIELARADSN I QALR HFAPVEDRLNA PG RD W RFP GDVWG W
 Sbjct: 74 ASLPQLWQLLIELARADSNITQALRGHFAFVEDRLNAPPGVDRITWHARFVAGDLVGCW 133

Query: 130 TEVGVFKIGDVITKVSAGQDG-FVINGTKFYSTGSI+PADWIDVYQADRAGDVI/VAVNA 188
 TEVGS+V+G+V+T+VS + DG +V+N+K+Y+STGS+F+DWID+YQAR D GADVIA +

-2592-

Sbjct: 134 TEVGSVRLGEVLTVSRKDDGRWVNGSKYISTGSLPSDWIDLNAQRDDTGADVIAAIRT 193

Query: 189 RHAGVRHSDWDGFGGQRTTG 208

GVR SDWDGFGGQRTTG

Sbjct: 194 DQFQVGRQSDWDGFGGQRTTG 213

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 2325

A DNA sequence (GBSx2468) was identified in *S.agalactiae* <SEQ ID 7139> which encodes the amino acid sequence <SEQ ID 7140>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1919(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 2326

A DNA sequence (GBSx2474) was identified in *S.agalactiae* <SEQ ID 7141> which encodes the amino acid sequence <SEQ ID 7142>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2978(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 2327

A DNA sequence (GBSx2476) was identified in *S.agalactiae* <SEQ ID 7143> which encodes the amino acid sequence <SEQ ID 7144>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5402(Affirmative) < succ>

-2593-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2328

- 10 A DNA sequence (GBSx2477) was identified in *S.agalactiae* <SEQ ID 7145> which encodes the amino acid sequence <SEQ ID 7146>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2755 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 190/247 (76%), Positives = 225/247 (90%)

- 25 Query: 1 MELGQLFHHLGSEITLQCRSERLLKEYDPEISESVKALISQINLVKGATFERVEQGSGE 60
 MELGQLFH+LGSE+TL+CRSERLLKEYDPEISESVK+L+EQGINLVKGAT+ER+EQ+G+
 Sbjct: 262 MELGQLFHHLGSEVTLQCRSERLLKEYDPEISESVKSLVSEQINLVKGATYERIEQNGD 321

- 30 Query: 61 IKRVYVTVNGSREVIBSDQLLVATGRKFWTDSIMLSAAGVETGKNNELINDFGQTSNEK 120
 IK+V+V VNG + +IE+DQLLVATGR FNT +LML AAGVE G EI+I+D+ +T+N +
 Sbjct: 322 IKKHVEVNGKKRIIEADQLLVATGRTPATATLNLRAAGVEIGSGEIIIDDYSRTTNR 381

- 35 Query: 121 TYAAGDVTLGPOFVYVAAYEGGIITDIAIGGLNKKIDLSVFAVFTFTNPTVATVGLTEEQ 180
 TYAAGDVTLGPOFVYVAAY+GG+ NAIGGLNKK++L VVP VTFF P +ATVGLTE+Q
 Sbjct: 382 TYAAGDVTLGPOFVYVAAYQGGVAAFNAGIGGLNKKLNLLEVVPGVFTTAPALATVGLTEQQ 441

- 40 Query: 181 AKKGYDVKTSVLPLQAVPRAIVNRETTGVFKLVADAECLKVLGVHIVSENAGDVIYAA 240
 AKK GY+VKTSVLPL AVFRA+VNRETTGVFKLVAD++T+KVLG H+V+ENAGDVIYAA+
 Sbjct: 442 AKKNGYEVKTSVLPLDQAVPRALVNRETTGVFKLVADSKTKMVLGARVVAENAGDVIYAA 501

- 45 Query: 241 LAVKFGSL 247
 LAVKFGSL
 Sbjct: 502 LAVKFGSL 508

There is also homology to SEQ ID 1820.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2329

- A DNA sequence (GBSx2478) was identified in *S.agalactiae* <SEQ ID 7147> which encodes the amino acid sequence <SEQ ID 7148>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2594-

```

bacterial cytoplasm --- Certainty=0.3642(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

- 5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2330

- 10 A DNA sequence (GBSx2479) was identified in *S.agalactiae* <SEQ ID 7149> which encodes the amino acid sequence <SEQ ID 7150>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

```

15

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1936(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

20

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2331

- 25 A DNA sequence (GBSx2480) was identified in *S.agalactiae* <SEQ ID 7151> which encodes the amino acid sequence <SEQ ID 7152>. This protein is predicted to be Nra. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

```

30

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1510(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35

A related GBS nucleic acid sequence <SEQ ID 9383> which encodes amino acid sequence <SEQ ID 9384> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7153> which encodes the amino acid sequence <SEQ ID 7154>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.64 Transmembrane 22 - 38 ( 22 - 38)

```

45

```

----- Final Results -----
bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-2595-

An alignment of the GAS and GBS proteins is shown below.

Identities = 42/157 (26%), Positives = 78/157 (48%), Gaps = 2/157 (1%)

```

5  Query: 71  LIASREFIDSQHFKDINAYFLRHFFICYCYFIPDFYFLNFSKLSY--SKDLYHLDKGLAD 128
      LLS  ++S  FK I  FR FI  +PD + + R  +K Y+ L  + +
      Sbjct: 8  LLNNIILNSLFPFKRILVFSFRLPISNLQVLLPDHPLHYLRQQRNKSFYNTLKATIVES 67

10  Query: 129  IFLKGGNLTFSKHETVLLTMQLSNLIZTFLAPLSVYVISNNIRLQTYQVMLANQYFESK 188
      + +G      + +L T+QL  L++T+L P+ VY+++++  L  L+ YF
      Sbjct: 68  WMSAGIVGKLPSYHLLLFITIQEEELACTYLPPPIFVYLLTNNTAALDMLNALSIFYPPPA 127

      Query: 189  IASFFVNYQTQIDRKLKKADIIAERNRYISSLKN 225
      IA  VN +      + +K  +IIA+R+Y++ +++
15  Sbjct: 128  IATVMPVNVVEIIPFKOIVKEQSVIADRYIANLIQH 164

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2332

- 20 A DNA sequence (GBSx2481) was identified in *S. agalactiae* <SEQ ID 7155> which encodes the amino acid sequence <SEQ ID 7156>. Analysis of this protein sequence reveals the following:

Possible site: 20
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1383 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2333

- 35 A DNA sequence (GBSx2482) was identified in *S. agalactiae* <SEQ ID 7157> which encodes the amino acid sequence <SEQ ID 7158>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

- 40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4145 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2596-

Example 2334

A DNA sequence (GBSx2484) was identified in *S.agalactiae* <SEQ ID 7159> which encodes the amino acid sequence <SEQ ID 7160>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.02    Transmembrane    34 - 50 { 34 - 50 }

----- Final Results -----
      bacterial membrane --- Certainty=0.1808 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2335

A DNA sequence (GBSx2485) was identified in *S.agalactiae* <SEQ ID 7161> which encodes the amino acid sequence <SEQ ID 7162>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3488 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GF:CAB52002 GB:AL109663 hypothetical protein [Streptomyces
coelicolor A3(2)]
Identities = 61/141 (43%), Positives = 86/141 (60%), Gaps = 2/141 (1%)

Query: 3 TYFDNPLKTNQAYADLGHGTAHLPIKPKTKVAIVTCMDSRSLHVAQALGLALGDNHILRNAG 62
T D ++ N+ YA + +P +VA+V CND+RL + ALGL LGD H +RNAG
35 Sbjet: 5 TVTDLRLVEANERYAAAFADPGMDARPVQRVAVVACMDARLDLHAALGLKLGDCHTIRNAG 64

Query: 63 GRVIDDVLRLSVISQQQLGTRTREVVLHHTDCGAQTFTNEAPAAQLQRLDGLVDMHGHEFLP 122
G VTDVV+RSL ISQ+ LGTR + ++HHT CG +T T E F L+ ++G
40 Sbjet: 65 GVVIDDVIRSLITISQALGTRSVALLIHTGCCGRTITRE-FRHDLELVG-QRPANAVEA 122

Query: 123 FNDIEESVREDVAKLIASPF 143
F D ++ VR+ ++ SPFL
Sbjet: 123 FRDADQDVQSIERVTSFPL 143

```

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6469> which encodes the amino acid sequence <SEQ ID 6470>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2295 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

- 55 An alignment of the GAS and GBS proteins is shown below.

-2597-

Identities = 109/146 (74%), Positives = 128/146 (87%)

Query: 1 MTTTFDNFLKTNQAYADLAGTNAHLPTKPKTKVAIVTCMSRLHVAQALGLALGDAILRN 60
 + +YF++F+ NQAY LKTAHL+KPKTKVAIVTCMSRLHVAQALGLALGDAILRN
 Sbjct: 1 LMSYFHFMAANQAYVALHGTNAHLPLKPKTKVAIVTCMSRLHVAQALGLALGDAILRN 60

Query: 61 AGGRVTDVLRSLVISQQQLGTREIVVLHHTDCGAGTFINBAFAQLQRDLGVDMHGHD 120
 AGGRVT+D++RSLVISQQC+STREIVVLHHTDCGAGTFINE FA + LGVD+ G DF
 Sbjct: 61 AGGRVTDMSRLVISQQQWSTREIVVLHHTDCGAGTFINBSFAKHTHEHLGVIVSGQDF 120

Query: 121 LFFNDIRSVREDNAKLHASPFLREE 146
 LFF D+E+SVRED+A+ AS + ++
 Sbjct: 121 LFFQDVEDSVREDNAKIRASSLSLSD 146

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2336

A DNA sequence (GBSx2486) was identified in *S.galactiae* <SEQ ID 7163> which encodes the amino acid sequence <SEQ ID 7164>. Analysis of this protein sequence reveals the following:

- 20 Possible site: 26
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.0932(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 30 >GP:AG08811 GB:AE004955 phosphoribosylaminoimidazole carboxylase,
 catalytic subunit [Pseudomonas aeruginosa]
 Identities = 20/27 (74%), Positives = 26/27 (96%)
- Query: 1 MFKHAEARGRGIKIILAGAGGAHLF 27
 +F++AEEA GRG+++ILAGAGGAHLF
 Sbjct: 46 LFQYAEAEARGGLEVIILAGAGGAHLF 72

There is also homology to SEQ ID 910:

- Identities = 27/27 (100%), Positives = 27/27 (100%)
- 40 Query: 1 MFKHAEARGRGIKIILAGAGGAHLF 27
 MFKHAEARGRGIKIILAGAGGAHLF
 Sbjct: 87 MFKHAEARGRGIKIILAGAGGAHLF 113

- 45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2337

A DNA sequence (GBSx2488) was identified in *S.galactiae* <SEQ ID 7165> which encodes the amino acid sequence <SEQ ID 7166>. Analysis of this protein sequence reveals the following:

- 50 Possible site: 43
 >>> Seems to have an uncleavable N-term signal seq
- | | | | |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -6.85 | Transmembrane | 58 - 74 (53 - 80) |
| INTEGRAL | Likelihood = -5.79 | Transmembrane | 103 - 119 (101 - 122) |
- Final Results -----
 bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2598-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

There is also homology to SEQ IDs 880 and 9278.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2338

A DNA sequence (GBSx2489) was identified in *S.galactiae* <SEQ ID 7167> which encodes the amino acid sequence <SEQ ID 7168>. This protein is predicted to be short chain alcohol dehydrogenase. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1742 (Affirmative)	< succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9357> which encodes amino acid sequence <SEQ ID 9358> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAD06605 GB:AB001530 putative oxidoreductase (Helicobacter pylori J99)
Identities = 68/94 (72%), Positives = 79/94 (83%)

Query: 4 IDLANNAGLALGLDKSYEADFGDHMTMINNVGLIYLTRCILPEKMWVEVNRGLIINLGS 63
ID LANNAGLALGL+K+YE + DW MI+TN+ GL+L+TR ILP M+R +G IINLGS
Sbjct: 76 IDALINNAGLALGLANKYCELDQMEVMDITNIGLLHLTRILFESMIEHDQSTIINLGS 135

Query: 64 XAGTIFVYGVANVYGASKAFVKQPSLNLRLADLAGT 97
AGT YPG NVYVGASKAFVKQPSLNLRLADLAGT
Sbjct: 136 IAGTYATPGGVNIVYGASKAFVKQPSLNLRLADLAGT 169

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7169> which encodes the amino acid sequence <SEQ ID 7170>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm ---	Certainty=0.0000 (Not Clear)	< succ>

A related sequence was also identified in GAS <SEQ ID 9121> which encodes the amino acid sequence <SEQ ID 9122>. Analysis of this protein sequence reveals the following:

Possible site: 12
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm ---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.